



A:Accession: A81919  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-447 <PAR>  
A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84035.1; PID:g737947  
A:Experimental source: serogroup A, strain 22491  
C:Genetics:  
A:Gene: nqrA; NMA0752  
C:Superfamily: Vibrio alginolyticus sodium-translocating NADH dehydrogenase (ubiquinone)  
C:Keywords: NAD; oxidoreductase

alignment\_scores:  
Quality: 2222.00 Length: 447  
Ratio: 4.982 Gaps: 0  
Percent Similarity: 99.776 Percent Identity: 96.197

alignment\_block:  
US-09-303-518d-131 x A81919

Align seg 1/1 to: A81919 from: 1 to: 447

1 ATGATTAAATCAAAAAGCTTAATCTGCCATCGCGGCGAGACCGGA 50  
1 MetilelysilelysylslyGlyLeuasnLeuProileAlaGlyArgProgl 17  
51 GCAAGTCATTATGACGGCGCGGCCCATATTACCGAAGTCGGTTCGCGG 100  
17 uGlnValileTyrAspGlyProValileThrGluValAlaLeuLeuGlyG 34  
101 AAGAATATGTCGGCATCGCGCCCTCGCATGAAATCAAGGAAGGTGAAGCC 150  
34 luGluTyrAlaGlyMetArgProSerMetlysVallysGluGlyAspAla 50  
151 GTCAAAAAGCGCAAGTCGTGTTTGAAGACAAAAGAAATCCGGCGCTAGT 200  
51 VallyslyslyGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67  
201 ATTACTGCGCGCGCTCAGGCAAAATCGCGCTATTACCGGCGGCAAA 250  
67 lPheThrAlaProValSerGlyLysileAlaAlaileHisArgGlyGluL 84  
251 AGCGCGTACTTCACTAGTCGTGATTGCCGTTGAAGGCAACACGAATC 300  
84 ysArgValLeuGlnSerValValileAlaValGluGlyAsnAspGluile 100  
301 GAGTTCGAAGCTACGTACCTGAGCGCTCGCAAAATTGAGCGGAA 350  
101 GluPheGluArgTyrAlaProGluAlaLeuAlaasnLeuSerGlyGluGl 117  
351 AGTGGCGCGCAACCTGATTCAATCAGGCTTATGACTCGCTTCGCAACC 400  
117 uValArgArgAsnLeuileGlnSerGlyLeuTyrThrAlaLeuArgThrA 134  
401 GTCGTTTACGAAATCCCTGCGCGTATGCGGAGCGGCTTCGCACTTC 450  
134 rgProPheSerLysileProAlaValAspAlaGluProPheAlailePhe 150  
451 GTCAATGCGATGGACACCAATTCGCTGGCTGCGGACCTTACGGTCAATC 500  
151 ValAsnAlaMetaspThrAsnProleuAlaAlaAspProValValVal 167  
501 CAAAGAAGCCCGCAAGACTTCAACACGCGGCTGTTGTTGATGAGCGGCC 550  
167 eLysGluAlaAlaGluAspPheArgGlyLeuLeuValLeuSerArgL 184  
551 TGACCGAAGCTTAAATCCATGTTGTAAAGCAGCAGCGGAGAGCTGCCG 600  
184 euThrGluArgLysileHisValCysLysAlaAlaGlyAlaAspValPro 200  
601 TCTGAAATGCTGCAATATCGAATACACATGAATTTGGCGCGCGCATCC 650  
201 SerGluAsnAlaAlaAsnileGluThrHisGluPheGlyProHisPr 217

651 TGCCGGCTTGAGTGGCAGCACATTCATTTTCATCGAGCCAGTCGCGCGA 700  
217 oAlaGlyLeuSerGlyThrHisileHisPheileGluProValGlyAla 234  
701 ATAAACCGTGTGGACCATCAATTAACAGACGTGATTCGTCAGCGCT 750  
234 snLysThrValTyrThrileAsnTyrGlnAspValileAlaileGlyArg 250  
751 TTGTTCTACAGCGCTCTGAATACCGGCGGTGGTTCGCTTCGCGCGG 800  
251 LeuPheAlaThrGlyArgLeuAsnThrGluArgValileAlaLeuGlyGl 267  
801 CCTGCAAGTCAACAAACCGCGCTCTTTCGCTACCGTTCGTCGCGAAG 850  
267 ySerGlnValAsnLysProArgLeuLeuArgThrValleuGlyAlaLys 284  
851 TGCTCAACTTACCGCGCGCAATTTGTTGACGGGACACCGCGGTGATT 900  
284 alSerGlnileThrAlaGlyGluLeuValAspAlaAspAsnArgValile 300  
901 TCCGTTTCGCTATTGAACGTTGCGATTTCACAGGCGCGCATATTATT 950  
301 SerGlySerValleuasnGlyAlaileThrGlnGlyAlaHisAspTyrLe 317  
951 GGGACGCTTACCAATCAGATTTCCGTTTATCGAAGAGCGCGCAGCAAG 1000  
317 uGlyArgTyrHisAsnGlnileSerValileGluGluGlyArgSerLys 334  
1001 AGCTGTTTCGCGTGGTTCGCGCGCGGACAAATATCTCATCCACGCGC 1050  
334 luLeuPheGlyTyrPValAlaProGlnProaspLysTyrSerileThrArg 350  
1051 ACCACTTCGCGCAATTTCTTAAAAAACAACCTTCAAGTTCACGACAGC 1100  
351 ThrThrLeuGlyHisPheLeuLysAsnLysLeuPheLysPheThrAl 367  
1101 GGTCAACGCGCGGACCGCGCATGGTACCGATCGGCACCTATTCAGCGCG 1150  
367 aValAsnGlyGlyAspArgAlaMetValProileGlyThrTyrGluArgV 384  
1151 TAATCCGCTTGACATCTCGCTACCTTCGCTTTTCGCGGATTAATGTC 1200  
384 alMetProLeuAspileLeuProThrLeuLeuLeuArgAspLeuileVal 400  
1201 GCGATACCGACGCGCGCGCGCTTTGGTTCGCTTGGAAATGGACGAGA 1250  
401 GlyAspThrAspSerAlaGlnAlaLeuGlyCysLeuGluLeuAspGluGl 417  
1251 AGACCTCGCTTGTGACGCTTCGCTCGCGGCAAAATACGAATACGCGCC 1300  
417 uAspLeuAlaLeuCysSerPheValCysProGlyLysTyrGluTyrGlyP 434  
1301 CGCTGTTGCGCAAGTTCGTTGAAACCATTTGAGAGGAAGGC 1341  
434 roleuLeuArgLysValLeuGluThrileGluLysGluGly 447

seq\_name: pir2:D81185  
seq\_documentation\_block:  
sodium-translocating NADH dehydrogenase (ubiquinone) (EC 1.6.5.-) alpha chain NMB0569  
C:Species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 02-Mar-2001  
C:Accession: D81185  
R:Tetelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, R.; Tetelin, H.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B. Hickey, E.K.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizzza, M. Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A:Reference number: A81000; MUID:20175755  
A:Accession: D81185  
A>Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-447 <TET>  
A:Cross-references: GB:AE002412; GB:AE002098; NID:g7225783; PIDN:AAF40997.1; PID:g722579  
A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB0569  
C:Superfamily: Vibrio alginolyticus sodium-translocating NADH dehydrogenase (ubiquinone)  
C:Keywords: oxidoreductase

alignment\_scores:  
Quality: 2221.00 Length: 447  
Ratio: 4.980 Gaps: 0  
Percent Similarity: 99.776 Percent Identity: 96.197  
alignment\_block:  
us-09-303-518d-131 x D81185 ..  
Align seg 1/1 to: D81185 from: 1 to: 447  
1 ATGATTAAATCAAAAAGGCTAAATCTGCCCATCGCGGAGACCGGA 50  
1 MetIleLysIleLysLysGlyLeuAsnLeuProIleAlaGlyArgProI 17  
51 GCAAGTCATTATGACGGCCCGCCCATACCGAAGTCGGTTCCTGGCG 100  
17 uGlnAlaValTyArgGlyProAlaIleThrGluValAlaLeuLeuGlyG 34  
101 AAGAATATCTGGCATGCGCCCTCGATGAAATCAAGAAAGTGAAGCC 150  
34 LuGluTyAlaGlyMetArgProSerMetLysValLysGluGlyAspAla 50  
151 GTCAAAAAGGCAAGTCTGTTTGAAGACAAAAAGAAATCCGGCGTAGT 200  
51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67  
201 ATTACTGCGCGGCTTCAGGCAAAATCCCGCTATTACCGTGGCGAAA 250  
67 lPheThrAlaProAlaSerGlyLysIleAlaAlaIleHisArgGlyGluL 84  
251 AGCGCGTACTTCAGTCAGTCTGATGCGGTTGAAGCAACGACCAATC 300  
84 ysArgValLeuGlnSerValValIleAlaValGluGlyAsnAspGluIle 100  
301 GAGTTCGAACGTACTGACGCTGGAAGCGCTGGCAAAATGACGAGCAAA 350  
101 GluPheGluArgTyAlaProGluAlaLeuAlaAsnLeuSerGlyGluG 117  
351 AGTGGCGCGCAACCTGATTCAATCAGGCTTATGGACTCGGCTTCGACCC 400  
117 uValArgArgAsnLeuIleGlnSerGlyLeuTrpThrAlaLeuArgThrA 134  
401 GTCGCTTCAGCAAAATCCCTGCTAGATGCGGAGCGCTTCGCCATCTTC 450  
134 rgProPheSerLysIleProAlaValAspAlaGluProPheAlaIlePhe 150  
451 GTCATGCGATGGACCAATCCGCTGGCTCGGACCCCTAGCGTCATCAT 500  
151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProThrValIleI 167  
501 CAAAGAACGCCCGCAAGACTTCAACGCGGCTGTTGGTATGAGCGGCC 550  
167 eLysGluAlaGluAspPheLysArgGlyLeuLeuValIleValIle 184  
551 TGACCGAACGTAAATCCATGTGTGTAAGCAGCAGCGCAGAGTGGCCG 600  
184 euThrGluArgLysIleHisValCysLysAlaAlaGlyAlaAspValPro 200  
601 TCTGAAATGCTGCAATATCGAACAACATCAATTTGGCGCGCCGATCC 650  
201 SerGluAsnAlaAlaAsnIleGluThrHisGluPheGlyGlyProHisPr 217  
651 TGGCGGCTTGAGTGGCAGCACATTCATTTCATCGAGCCAGTCGGCGCA 700

|||||  
217 oAlaGlyLeuSerGlyThrHisIleHisPheIleGluProValGlyAlaA 234  
701 ATAAACCCGTGGACCATCAATTATCAAGAGCTGATTGCTATCGGAGCT 750  
|||||  
234 snLysThrValTrpThrIleAsnTyArgInAspValIleThrIleGlyArg 250  
751 TTGTTCTGAACAGCGCTCTGAATACCGAGCGCTGGTGTCTGGCGGG 800  
251 LeuPheAlaThrGlyArgLeuAsnThrGluArgValIleAlaLeuGlyG 267  
801 CCTGCAAGTCAACAACCGCCCTCTCGGTACCGCTTTGGTGGTGGAGG 850  
267 ySerGlnValAsnLysProArgLeuLeuArgThrValLeuGlyAlaLysV 284  
851 TGTCTCAACTTACCGCGCGGAATTGTTGACGCGGACCAACCGCTGATT 900  
284 alSerGlnIleThrAlaGlyGluLeuValAspThrAspAsnArgValIle 300  
901 TCCGTTTCGGTATTGAACGGTTCGATTGCACAGCGCGCATGATTATT 950  
301 SerGlySerValLeuAsnGlyAlaIleThrGlnGlyAlaHisAspTyLe 317  
951 GGGACGCTACCAATCAGATTCCGTTATCCAGAAGAGCGCGCAGCAAG 1000  
317 uGlyArgTyHisAsnGlnIleSerValIleGluGluGlyArgSerLysG 334  
1001 AGCTGTTGCGTGGTTCGCGCGCAGCGGACAAATACTCCATCACGCGC 1050  
334 LuLeuPheGlyTrpValAlaProGlnProAspLysTySerIleThrArg 350  
1051 ACCACTCTCGGCCATTTCCTAAAAACAACTCTCAAGTTCACGACAGC 1100  
351 ThrThrLeuGlyHisPheLeuLysAsnLysLeuPheLysPheAsnThrAl 367  
1101 CGTCAACGCGCGCACCGCCATGTTACCGATCGGCGCCTTATGACGCG 1150  
367 aValAsnGlyGlyAspArgAlaMetValProIleGlyThrTyGluArgV 384  
1151 TAATGCGCTGGACATCCTGCTACCTGCTTTGCGCGATTTAATCGTC 1200  
384 alMetProLeuAspIleLeuProThrLeuLeuLeuArgAspLeuIleVal 400  
1201 GGCATACCGCACAGCGCGAGCTTGGGTGCTTGGATTGGACGAAGA 1250  
401 GlyAspThrAspSerAlaGlnAlaLeuGlyCysLeuGluLeuAspGluG 417  
1251 AGACCTCGCTTTGTGACAGCTTCGCTGCCGCGGCAATACGAATACGCGC 1300  
417 uAspLeuAlaLeuCysSerPheValCysProGlyLysTyThrGluTyGlyP 434  
1301 CGCTGTGCGCAAGTGTGCGAAACCATTTGAGAAGGAAGGC 1341  
434 roLeuLeuArgLysValLeuGluThrIleGluLysGluGly 447  
seq\_name: pir2:164002

seq\_documentation\_block:  
Sodium-translocating NADH dehydrogenase (ubiquinone) (EC 1.6.5.-) alpha chain HI0164  
C:Species: Haemophilus influenzae  
C:Date: 18-Aug-1995 #sequence\_revision 02-Sep-2000 #text\_change 02-Mar-2001  
C:Accession: I64002; A64003  
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman  
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter  
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630  
A:Accession: 164002  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-447 <TIGR>

A:Cross-references: GB:L42023; TIGR:HI0164; GB:U32702; NID:gl573118; PIDN:AAC21836.1; PI  
A:Note: the sequence is revised in GenBank entry U32702, PID:gl573122  
C:Superfamily: Vibrio alginolyticus sodium-translocating NADH dehydrogenase (ubiquinone)  
C:Keywords: oxidoreductase

alignment\_scores:  
Quality: 1646.00 Length: 448  
Ratio: 4.146 Gaps: 2  
Percent Similarity: 88.616 Percent Identity: 70.089  
alignment\_block:  
US-09-303-518d-131 x I64002  
Align seg 1/1 to: I64002 from: 1 to: 447

1 ATGATTAAATCAAAAGAGTCTAAATCTGCCCATCGCGGCGAGACCGGA 50  
1 MetileThrIleLysGlyLeuAspLeuProIleAlaGlyLysProAl 17  
51 GCAAGTCATTATGACGGCCCGCCATTACCGAAGTCGCGTTCGTTGGCG 100  
17 aGlnValIleHisSerGlyAsnAlaValAsnGlnValAlaIleLeuGly 34  
101 AGAATATGTCGCATCGCCCGCCCTCGATGAAATCAAGGAGGTGAAGCC 150  
34 luGluTyrValGlyMetArgProSerMetLysValArgGluGlyAspVal 50  
151 GTCAAAAAGCGCAAGTCGTTTGAAGACAAAAGAAATCCGGCGTAGT 200  
51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValI 67  
201 ATTACTGCGCGGCTTCAGGCAAAATCCCGCTATTCACCGTGGCGAAA 250  
67 ePheThrAlaProIleSerGlyThrIleThrAlaIleAsnArgGlyGlu 84  
251 AGCGGTACTTCAGTCAGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 300  
84 ysArgValLeuGlnSerValValIleAsnValGluGlyAspGluLysI 100  
301 GAGTTCGAAGCTACGTCACCTGAAGCCCTGGCAAAATGAGCAGCGAAA 350  
101 ThrPheAlaLysTyrSerThrGluGlnLeuAsnThrLeuSerSerGluG 117  
351 AGTCGCGCGCAACCTGATCAATCAGCTGATGACTGCGCTTCGCAACC 400  
117 nValLysGlnAsnLeuIleGluSerGlyLeuThrAlaLeuArgThrA 134  
401 GTCGCTTCAGCAAAATCCCTGCCGTAGATGCCGAGCGCTTCGCATCTC 450  
134 rgProPheSerLysValProSerIleGluSerGluAlaSerSerIlePhe 150  
451 GTCATATGGATGGACCAATCCGCTGCTGCCGACCCCTACGGTCATCAT 500  
151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProSerValVal 167  
501 CAAGAAGCCCGGAGACTTCAACAGCGGCTGTGTTGATTGAGCGGCC 550  
167 uLysGluTyrSerGlnAspPheThrAsnGlyLeuThrValLeuSerArgL 184  
551 TGACCGAAGCTAAA...ATCATGTGTGTAAGCAGCAGCGCGCAGACGTG 597  
184 euPheProSerLysProLeuHisLeuCysLysAlaGlyAspSerAsnIle 200  
598 CGGTCGTAATGTCGCCAATATCAACACATGAAATTTGGCGCCCGCA 647  
201 ProThrAlaAspLeuGluAsnLeuGlnIleHisaspPheThrGlyValH 217  
648 TCCTCGCGGCTGAGTCGACGACCATTCATTCATCGACGACGTCGGCG 697  
217 sProIleGlyLeuValGlyThrHisIleHisPheIleAspProValGlyI 234  
698 CGAATAAAACCGGTGGGACCATCAATATCAAGACGTGATTCCTATCGGA 747

234 leGlnLysThrValTrpHisIleAsnTyrGlnAspValIleAlaValGly 250  
748 CGTTTGTTCGTACAGCCGCTCTGATACCGAGCGGCTGTTGCCTTGGG 797  
251 LysLeuPheThrThrGlyGluLeuTyrSerGluArgValIleSerLeuAl 267  
798 CGGCTGCAAGTCAACAAACCGGCTCTTTCGTACCGTTTGGTGGCGGA 847  
267 aglyProGlnValLysGluProArgLeuValArgThrThrIleGlyAlaA 284  
848 AGTGTCTCAACTTACCGCGCGGCAATTTGTTGACGCGGACAAACCGCG 897  
284 snLeuSerGlnLeuThrGlnAsnGluLeuSerAlaGlyLysAsnArgVal 300  
898 ATTTCCGGTTCGGTATTTGAACGGTTCGATTCACAAAGCGCGCATGATTA 947  
301 IleSerGlySerValLeuCysGlyGlnIleAlaLysAspSerHisAspTy 317  
948 TTGGGACGCTACCAATCAGATTTCGTTATCGAAGAGCGCGGACGA 997  
317 rLeuGlyArgTyrAlaLeuGlnValSerValIleAlaGluGlyAsnGluL 334  
998 AAGAGCTGTCGCTGCGTTCGCGCGGAGCGGACGACAAATCTCTCATCAG 1047  
334 ysGluPhePheGlyTrpIleMetProGlnAlaAsnLysTyrSerValThr 350  
1048 CGCACCTCTCGGCAATTCCTTAAACAACTCTTCAAGTTTCAGGAC 1097  
351 ArgThrValLeuGlyHisPheSerLys...LysLeuPheAsnPheThrTh 366  
1098 AGCGCTCAACGCGCGCGCATGTCACCGATCGGACGCGCATTCATGAGC 1147  
366 rSerGluAsnGlyGlyGluArgAlaMetValProIleGlySerTyrGluA 383  
1148 GCGTAATGCGCTGACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1197  
383 rgValMetProLeuAspIleLeuProThrLeuLeuLeuArgAspLeuIle 399  
1198 GTCGCGATACCGACGCGCGCGGCTTTCGCTGCTGCTGCTGCTGCTGCTG 1247  
400 ValGlyAspThrAspGlyAlaGlnGluLeuGlyCysLeuGluLeuAspG 416  
1248 AGAAGACCTCGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1297  
416 uGluAspLeuAlaLeuCysSerPheValCysProGlyLysTyrGluTyrG 433  
1298 GCGCGCTGTCGCAAGTCTGGAACCATTCGAGAGGAAGGC 1341  
433 lySerIleLeuArgGlnValLeuAspLysIleGluLysGluGly 447

seq\_name: pir2:AG0393

seq\_documentation\_block:  
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain A [imported] - Yersinia pestis (st  
C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AG0393  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.  
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrel  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-447 <KUP>  
A:Cross-references: GB:AL590842; PIDN:CAC92475.1; PID:gl5981176; GSPDB:GN00175  
C:Genetics:  
C:Superfamily: Vibrio alginolyticus sodium-translocating NADH dehydrogenase (ubiquino  
C:Keywords: Oxidoreductase

alignment\_scores:  
Quality: 1600.00 Length: 447  
Ratio: 4.071 Gaps: 0  
Percent Similarity: 87.919 Percent Identity: 69.128

alignment\_block:  
US-09-303-518D-131 x AG0393 ..  
Align seg 1/1 to: AG0393 from: 1 to: 447

```
1 ATGATTAAATCAAAAGGCTTAATCTGCCATCGCGCGAGACCGGA 50
1 MetileLysileLysLysGlyLeuAspLeuProileAlaGlyAlaProVa 17
51 GGAAGTCATTATGAGCGCGCGCCATACCAAGTCCGGTTCGTTGGCG 100
17 lGlnThrileGlnAspGlyProAlaileHisValAlaLeuLeuGlyG 34
101 AAGAATATGTCGGCATCGCCCTCGATGAAATCAAGGAAGTGAAGCC 150
34 luGluTyValGlyMetArgProSerMetLeuValGlnGluGlyAspGln 50
151 GTCAAAAGGCCAAGTCTGTTTGAAGACAAAAGATCCGGCGTAGT 200
51 ValLysLysGlyGlnAlaLeuPheGluAspLysLysAsnProGlyValle 67
201 ATTTACTGCGCGGCTTCAGGCAAAATCGCGCTATTACCGTGGCGAAA 250
67 uphetrAlaProAlaSerGlyLysileSerAlaileAsnArgGlyGluA 84
251 AGCGCGTACTTCAAGTCACTGCTGATGCGGTTGAAGCAACGAGAAATC 300
84 rgArgValLeuGlnSerValVallieGluValGluGlyAspGluGlnle 100
301 GAGTTCGAACCTACGTACCTGAGCGCTGCAAAATTTGACGACGAAA 350
101 PropheGluHisTyAlaAlaGluLeuLeuAsnGlnLeuSerAspGluG 117
351 AGTGGCGCGCAACCTGATTCAATCAGGCTTATGACTGCGCTTCGACCC 400
117 nValGlnHisHisLeuLeuThrSerGlyLeuTyPrThrAlaLeuArgthr 134
401 GTCGTTTCAGCAAAATCCCTCCCTAGATGCGGAGCGCTTCGCCATTC 450
134 rgProPheSerLysThrProValProGlySerArgProArgAlailephe 150
451 GTCATGCGATGGACACCAATCCGCTGCTGCGCGACCCCTACGGTCATCAT 500
151 lIleSerAlaMetAspThrGlnProLeuAlaAlaAspProGlnValileil 167
501 CAAGAAGCGCGCAAGACTTCAACGCGCGCTGTTGTTGATGAGCGGCC 550
167 eAlaThrGluSerGluAlaPheAsnHisGlyLeuThrValLeuThrArgL 184
551 TGACCGAAGCTTAATCCTGCTGTAAACGACGACGCGCGAGAGCTGCCG 600
184 euthrAspGlyLysValHisValCysHisAlaAlaGlyGlnAlaValThr 200
601 TCTGAAATGCTGCCAATATCAATCAACACATGAATTTGGCGCGCCGATCC 650
201 ArgHisThrAsnThrGlnValThrTyAsnGluPheSerGlyProHisPr 217
651 TGCGCGCTTGAGGCGACGACATTCATTCATCGAGCCAGTCCGGCGCA 700
217 oAlaGlyLeuValGlyThrHisleHisPheLeuGluProValSerGln 234
701 ATAAACCGTGTGCACCATCAATATCAACACGCTGATTCGTCGACGT 750
234 hrLysMetValTphHisValGlyTyGlnAspVallieAlaileGlyLys 250
751 TTGTTCTAAACGCGCTCTGAATACCGCGCGCTGTTGCGTTCGGCGG 800
```

```
251 LeuPheThrArgGlyGluLeuCysThrAspArgileValAlaLeuAlaG 267
801 CCTGCAAGTCAACAAACGCGCTCTTGGTACCGTTTGGGTGCGAAGG 850
267 yProGlnValAsnGlnProileLeuLeuLeuArgThrArgLeuGlyAlaSerL 284
851 TGTCTCAACTTACCGCGCGGAATTTGGTTCAGCGGACAAACCGCGTATT 900
284 euSerGluLeuThrAlaGlyLysLeuLysGluGlyAspAsnArgilele 300
901 TCCGGTTCGGTATTCAACCGTGCATTCACAAAGCGCGCATGATTATT 950
301 SerGlySerValLeuSerGlyThrAlaPheSerAlaThrHisGlyTyLe 317
951 GGGACGCTTACCACCAATCAGATTTCGGTTATCGAAGAGCGCGCAGCAAG 1000
317 uGlyArgPheHisGlnValSerValileArgGluGlyArgGluLysG 334
1001 AGCTGTTGCGGCTGGTTCGCGCGAGCGCGCAAAATCTCCATCACGCGC 1050
334 luLeuPheGlyTrpValMetProGlyArgAspLysTySerileThrArg 350
1051 ACCACTCTCGGCCATTTCTTAAACAAACTCTTCAAGTTCAGCAGC 1100
351 ThrThrLeuGlyHisPhePheLysArgLysLeuPheAlaPheSerThrAs 367
1101 CGTCAACGGCGGCGACCGCGCCATGTGACCGATCGGACCTATGAGCGG 1150
367 pMetHisGlyGlyGluArgAlaMetValProileGlyAsnTyGluArgv 384
1151 TAATGCGGTTGGACATCTCCCTACCTGCTGTTTGGCGGATTTAATCGTC 1200
384 alMetProLeuAspilleLeuAlaThrHisLeuLeuArgAspLeuLeuAla 400
1201 GGCATATACCGACAGCGCGAGGCTTGGTTCGTTGGAATTCGACCAAGA 1250
401 GlyAspThrAspSerAlaGlnAlaLeuGlyCysLeuGluLeuAspGluG 417
1251 AGACCTCGCTTGTGAGCTTCGTCGCGCGGCAAAATACGAAATACGCGC 1300
417 uAspLeuAlaLeuCysThrPheValCysProGlyLysTyGlyTyGlyP 434
1301 CGCTGTTGGCAAGTCTGGAACCATTCAGAAAGCGC 1341
434 roValLeuArgAspilleLeuThrGlnileGluGluGly 447
```

seq\_name: pir2:S51015

seq\_documentation\_block:

sodium-translocating NADH dehydrogenase (ubiquinone) (EC 1.6.5.-) alpha chain ngra [v  
N:Alternate names: ngra protein  
C:Species: Vibrio alginolyticus  
C:date: 10-Apr-1996 #sequence\_revision 19-Apr-1996 #text\_change 16-Feb-2001  
C:Accession: S51015; S66365  
R:Beattie, P.; Tan, K.; Bourne, R.M.; Leach, D.; Rich, P.R.; Ward, F.B.  
FBS Lett. 356, 333-338, 1994  
A:Title: Cloning and sequencing of four structural genes for the Na(+)-translocating  
A:Reference number: S51013; MUID:95104445  
A:Accession: S51015  
A:Molecule type: DNA  
A:Residues: 1-446 <BEA>  
A:Cross-references: EMBL:Z37111; NID:9663268; PIDN:CAA85476.1; PID:g663270  
A:Accession: S66365  
A:Molecule type: protein  
A:Residues: 1-10 <BEW>  
C:Genetics:  
A:Gene: ngra

C:Superfamily: Vibrio alginolyticus sodium-translocating NADH dehydrogenase (ubiquino  
C:Keywords: FMN; oxidoreductase

alignment\_scores:

Quality: 1488.50 Length: 447  
Ratio: 3.856 Gaps: 1  
Percent Similarity: 86.353 Percent identity: 63.758

alignment\_block:  
US-09-303-518D-131 x S51015 ..

Align seg 1/1 to: S51015 from: 1 to: 446

1 ATGATTAATAAATAAAGTCTAATCTGCCATCGCGGCGAGACCGA 50  
|||||  
1 MetIleThrIleLysLysGlyLeuAspLeuProIleAlaGlyThrProSe 17  
51 GCAAGTCATTATGACGCGCGCCGACATACCGAAGTCGCGTTCGGG 100  
:|||||  
17 rGlnValIleAsnAspGlyLysThrIleLysLysValAlaLeuGlyG 34  
101 AGAATATGTCGCGATCGCCCTCGATCAATAAATCAAGGAAGTGAAGCC 150  
|||||  
34 LuGlyThrValGlyMetArgProThrMetHisValArgValGlyAspGlu 50  
151 GTCAAAAAGGCAAGTGTGTGTAAGACAAAGAAAGATCCGGCGTAGT 200  
|||||  
51 ValLysLysAlaGlnValLeuPheGluAspLysLysAsnProGlyVally 67  
201 ATTACTGCGCGCTCAGGCAAAATCGCGCTATTACCGTGGCGAAA 250  
|||||  
67 sPheThrAlaProAlaAlaGlyLysValIleGluValAsnArgGlyAlaL 84  
251 AGCGCGTACTTCACTCAGTCGTGATTCGCGTTGAAGGCAACGACGAATC 300  
|||||  
84 ysargValLeuGlnSerValValIleGluValAlaGlyGluGlnVal 100  
301 GAGTTCGACGCTACGTACCTGAGCGCTGCGCAAAATGAGCAGCGAAA 350  
|||||  
101 ThrPheAspLysPheGluAlaAlaGlnLeuSerGlyLeuAspArgGluVa 117  
351 AGTCGCGCGCAACTGATTCATCAATCAGCTTATGACTGCGCTTCGCACCC 400  
|||||  
117 lIleLysThrGlnLeuValAspSerGlyLeuThrAlaLeuArgThra 134  
401 GTCCGTTACGAAATCCCTCGCGTAGATCCGAGCGCTTCGCACATCTC 450  
|||||  
134 rgProPheSerLysValProAlaIleGluSerThrLysAlaIlePhe 150  
451 GTCATCGGATGACACCAATCCGCTGCGTGCAGCCCTACGTCATCAT 500  
|||||  
151 ValThrAlaMetAspThrAsnProLeuAlaAlaLysProGluLeuIle 167  
501 CAAGAAGCGCGCGCAAGACTTCAACCGCGCTGTTGTTATGAGCGGCC 550  
|||||  
167 eAsnGluGlnGluAlaPheIleAlaGlyLeuAspIleLeuSerAlaL 184  
551 TGACCGAAGCTTAATCCATGTGTGTAAGCAGCAGCGCGAGCGTCCG 600  
|||||  
184 euThrGluGlyLysValThrValCysLys...SerGlyThrSerLeuPro 199  
601 TCTGAATATGTCGAATATGCAATACATGAATTTGGCGCGCGCATCC 650  
:|||||  
200 ArgSerSerGlnSerAsnValGluGluHisValPheAspGlyProHisPr 216  
651 TGCGCGGTTGAGTGGCGCAGCATTCATTCATCGAGCCAGTCGCGCGA 700  
|||||  
216 oAlaGlyLeuAlaGlyThrHisMetHisPheLeuThrProValAsnAlaG 233  
701 ATAAACCGTGTGACCATCAATATCAAGACGTGATTGTCGAGCGT 750  
:|||||  
233 luAsnValAlaTrpSerIleAsnThrGlnAspValIleAlaPheGlyLys 249  
751 TTGTTCTGTAACAGCGCTCAATACCGAGCGCTGTTGCTTGGCGCG 800  
|||||  
250 LeuPheLeuThrGlyGluLeuThrThrAspArgValValSerLeuAlaGl 266

alignment\_scores:  
Quality: 1476.50 Length: 447  
Ratio: 3.845 Gaps: 1

seq\_name: pir2:G82094  
seq\_documentation\_block:  
sodium-translocating NADH dehydrogenase (ubiquinone) (EC 1.6.5.-) nqrA VC2295 [import  
C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 16-Feb-2001  
C:Accession: G82094  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.  
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers  
L, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833  
A:Accession: G82094  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-462 <HEI>  
A:Cross-references: GB:AF004300; GB:AE003852; NID:99656850; PIDN:AAF95439.1; GSPDB:GN  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VC2295  
A:Map position: 1  
C:Superfamily: Vibrio alginolyticus sodium-translocating NADH dehydrogenase (ubiquino  
C:Keywords: Oxidoreductase



Percent Similarity: 85.906 Percent Identity: 62.416

alignment\_block:

US-09-303-518D-131 x G82094

Align seg 1/1 to: G82094 from: 1 to: 462

```
1  ATGATTAAATCAAAAAGGCTTAATCTGCCATCGCGGCGGACACCGGA 50
   |||||
17 MetileThrIleLysLysGlyLeuAspLeuProIleAlaGlyThrProSe 33
   |||||
51 GCAGTCAATTTATGACGGCGCGCCGCAATACCGAAGTCGCTTGGTGGCG 100
   |||||
33 rGlnValIleSerAspGlyLysAlaIleLysLysValAlaLeuLeuGlyG 50
   |||||
101 AAGATATATCGCGCATCGCCCTCGATGAAATCAAGAGGTCGAAGCC 150
   |||||
50 luGluTyValGlyMetArgProThrMetHisValArgValGlyAspGlu 66
   |||||
151 GTCAAAAAGGCCAAGTCTGTTTGAAGACAAAGATCCGGCGGTAGT 200
   |||||
67 ValLysLysAlaGlnIleLeuPheGluAspLysLysAsnProGlyValLy 83
   |||||
201 ATTACTCGCGCGCTTCAGCAAAATCCGCGCTATTACCGTGGCGAAA 250
   |||||
83 sPheThrSerProValSerGlyLysValValGluIleAsnArgGlyAlaL 100
   |||||
251 AGCGGTACTTTCAGTCACTGCTGATTCGCGTTCGAGGCAAGCAAGTATC 300
   |||||
100 ysArgValLeuGlnSerValValIleGluValAlaGlyAspAspGlnVal 116
   |||||
301 GAGTTCGACGCTAGTACCTGACGCGTTCGAGGCGTTCGAGGCGGAAA 350
   |||||
117 ThrPheAsnGlyPheGluAlaAsnGlnLeuAlaSerLeuAsnArgAspAl 133
   |||||
351 AGTGGCGCGCAACCTGATTCATCAATCAGCTGATGACCTGCGCTTCGAC 400
   |||||
133 aIleLysThrGlnLeuValGluSerGlyLeuThrAlaPheArgThrA 150
   |||||
401 GTCCGTTTCAGCAAAATCCGCGCTAGATCGCGAGCGTTCGCGCATCTTC 450
   |||||
150 rgProPheSerLysValProAlaIleAspSerThrSerGluAlaIlePhe 166
   |||||
451 GTCATGCGATGGACACCAATCCGCTGCGTTCGCGACCTACGTCATCAT 500
   |||||
167 ValThrAlaMetAspThrAsnProLeuAlaAlaGluProThrValValIle 183
   |||||
501 CAAGAAGCGCGCGGACGACTTCAACGCGCGCTGTTGGTATTGACGCGCC 550
   |||||
183 eAsnGluGlnSerGluAlaPheValAlaGlyLeuAspValLeuSerAlaL 200
   |||||
551 TGACCGAAGCTAAATCCATGCTGTGTAAACGACGACGCGGCGGCGCG 600
   |||||
200 euThrThrGlyLysValTyValCysLys...LysGlyThrSerLeuPro 215
   |||||
601 TCTGAAATGCTGCAATATCGAAACACATGATTTGGCGCGCCGATCC 650
   |||||
216 ArgSerGlnGlnProAsnValGluGluHisValPheAspGlyProHisPr 232
   |||||
651 TGCGCGCTTACGTCGACGACCATTCATTTCATCGAGCGGCGCGCGA 700
   |||||
232 oAlaGlyLeuAlaGlyThrHisMetHisPheLeuTyProValSerAlaA 249
   |||||
701 ATAAACCGGTGGACATCAATATCAAGACGTGATTCCTATCGGACGT 750
   |||||
249 sPHisValAlaIleTrpSerIleAsnTyrgLnaAspValIleAlaValGlyGln 265
   |||||
751 TTGTTCTGAACAGCGCTCTGAATACCGAGCGCGTGTTCCTTGGCGCG 800
   |||||
266 LeuPheLeuThrGlyGluLeuTyThrGlnArgValValSerLeuAlaG 282
   |||||
801 CCTGCAAGTCAACAAACCGCGCTCTGCGTACCGGTTTGGGTGCGAAGG 850
   |||||
```

```
282 yProValValAsnLysProArgLeuValArgThrValMetGlyAlaSerL 299
851 TGTCTCAACTTACCGCGCGCGAATGTTGTTGCGCGGACAAACCGGTGATT 900
   |||||
299 euGluGlnLeuValAspSerGluIleMetProGlyGluValArgIleLe 315
   |||||
901 TCGGTTTCGGTATTGACGCGTGCATGTCACAAAGCGCGCGCATGATTAT 950
   |||||
316 SerGlySerValLeuSerGlyThrLysAlaThrGlyProHisAlaTyLe 332
   |||||
951 GGGAGCTACCAACAATCAGATTTCGTTATCGAAGAGCGCGCGCAAG 1000
   |||||
332 uGlyArgTyHisLeuGlnValSerValLeuArgGluGlyArgAspLysG 349
   |||||
1001 AGCTGTTTCGGTTCGGTTCGGCGCGGACGCGGACAAATCTCCATCACGCG 1050
   |||||
349 luLeuPheGlyTrpAlaMetProGlyLysAsnLysPheSerValThrArg 365
   |||||
1051 ACCACTTCGCGCATTTCTCTAAAACAACTCTCAAGTTCACGACAGC 1100
   |||||
366 SerPheLeuGlyHisLeuPheLysGlyGlnValTyAsnMetThrThrTh 382
   |||||
1101 CGTCAACGCGCGCGGACGCGCATGTTACCGTTCGCGCATTCATGAGCGG 1150
   |||||
382 rThrAsnGlySerAspArgSerMetValProIleGlyAsnTyrgLulysV 399
   |||||
1151 TAATCCGCTTGGACATCTCTACCTTGTCTTTCGCGCATTTAATCGTC 1200
   |||||
399 alMetProLeuAspMetGluProThrLeuLeuLeuArgAspLeuCysAla 415
   |||||
1201 GCGGATACCGACGCGCGCGCGCTTTCGGTTCGTTGGAATTCGACGAAGA 1250
   |||||
416 GlyAspSerAspSerAlaValArgLeuGlyAlaLeuGluLeuAspGluG 432
   |||||
1251 AGACTCGCTTTCGACGCTTCGCTCCCGCGGCAATACGAATACGCGCC 1300
   |||||
432 uAspLeuAlaLeuCysThrPheValCysProGlyLysTyrgLulysGlyG 449
   |||||
1301 CGCTCTTCGCGCAAGTCTCGAAACCATTCGAGAGGAGGCG 1341
   |||||
449 lnLeuLeuArgGluCysLeuAspLysIleGluLysGluGly 462
```

seq\_name: pir2:H83272

seq\_documentation\_block:

sodium-translocating NADH dehydrogenase (ubiquinone) (EC 1.6.5.-) nqr1 chain PA2999  
C:Species: Pseudomonas aeruginosa  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 16-Feb-2001  
C:Accession: H83272  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: AB2950; MUID:20437337

A:Accession: H83272

A:Molecule type: DNA

A:Residues: 1-445 <STO>

A:Cross-references: GB:AE004724; GB:AE004091; NID:g9949083; PIDN:AAG06387.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: nqrA; PA2999

C:Superfamily: Vibrio alginolyticus sodium-translocating NADH dehydrogenase (ubiquino

C:Keywords: Oxidoreductase

alignment\_scores:

Quality: 1358.00

Ratio: 3.641

Percent similarity: 83.445

Percent identity: 59.060

alignment\_block:

```
US-09-303-518D-131 x H83272 ..
Align seg 1/1 to: H83272 from: 1 to: 445

1 ATGATTAAATCAAAAAAGGCTTAATTCGCCATCGCGGCGACACGGGA 50
1 MetIleLysIleValGlyLeuSerGlyAlaProGlu 17
51 GCAAGTCATTTATGACGGCGCGCCATACCGAAAGTCGCTGGTGGCG 100
17 uGlnArgIleGluAlaAArgProValArgSerValAlaLeuIleGlyP 34
101 AGAATATCTCGCATCGCCCTCGATCAAAATCAAGGAAGTCAAGCC 150
34 heAspTyrHisGlyMetLysProThrMetAlaValGlnValGlyAspArg 50
151 GTCAAAAAGGCCAAGTCGCTGTTTGAAGACAAAAAGAAATCCGGCGTAGT 200
51 ValLysLeuGlyGlnValLeuPheThrAspLysLysAsnProSerValSe 67
201 ATTTACTGCGCGCTTCAGGCAAAATCGCCGCTATTACCGTGGCGAAA 250
67 rTyrThrAlaProGlyAlaGlyValValSerAlaIleHisArgGlyGluL 84
251 AGCGGTACTTTCAGTCAGTCGCTGATTCGCGTTGAAGGCAACGACGAAATC 300
84 ysArgValLeuGlnSerValValIleAspLeuAspGlyAspGluGlnLeu 100
301 GAGTTCGAAGCGTACCTGAGCGCTGCGCAAAATTTGAGCAGCGAAA 350
101 GluPheAlaArgTyrProAlaAspLysLeuAlaThrLeuSerAlaGluG 117
351 AGTGGCGCGCAACTGATTCATCAGGCTTATGAGCTGCGCTTCGACCC 400
117 nValArgAspAsnLeuGlnSerGlyLeuThrAlaLeuArgThrA 134
401 GTCGTTTCAGCAAAATCCCTGCGGTAGATGCGGAGCGCTTCGCCATCTTC 450
134 rgProPheSerLysValProAspProGluSerSerProSerIlePhe 150
451 GTCATTCGATGACCAATCGCTGCGTGGCGGACCTACGCTCATCAT 500
151 ValThrAlaIleAspThrGlnProLeuAlaAlaAspProGlnValValI 167
501 CAAGAAGCGCGCGAAGCTTCAACGCGCGCTGTTGTTGATGAGCGCG 550
167 eAlaGluGlnGlyGluAlaPheGlnAlaGlyLeuThrValLeuGlyArgL 184
551 TGACCGAAGCTTAAATCCATGTGTGTAAAGCAGCAGCGCGCAGCGCGG 600
184 euAla.....ArgValPheLeuCysLysAlaGluGlyValSerLeuPro 198
601 TCTGAAATGCTGCCAATATCGAAACACATGAATTTGGCGGCGCGCATCC 650
199 GlyGluAlaLeuSerGlyValThrAlaGlnAlaPheSerGlyProHisr 215
651 TGCGGCTTGTAGTCGACGCGCATTTTCATCAGCGCCAGCTCGCGCGGA 700
215 oAlaGlyLeuProGlyThrHisIleHisPheLeuAspProValGlyAlaG 232
701 ATAAACCGGTGACCATCAATTTATCAACAGCTGATGCTATCGGACGT 750
232 lyLysSerValTyrAsnLeuAsnTyrGlnAspValIleAlaIleGlyLys 248
751 TTGTTCGTAACAGCGCTGCTGAATACGAGCGGTGCTTGCCTTGGCGGG 800
249 LeuPheThrThrGlyGlnLeuThrPheGluArgValIleAlaLeuAlaG 265
801 CCTGCAAGTCAACAAACCGCGCTCTTCGCTACCGTTTGGTGGCGAAGG 850
265 yProValValGluLysProArgLeuValArgThrArgLeuGlyAlaAsnL 282
851 TGCTCTCACTTACCGCGCGCAATTTGTTGACGCGGACACCGCGTGATT 900
```

```
seq_name: pir2:C81751
seq_documentation_block:
probable sodium-translocating NADH dehydrogenase (ubiquinone) (EC 1.6.5.-) alpha chain
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 16-Feb-2001
C:Accession: C81751
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A:Reference number: AB1500; MUID:20150255
A:Accession: C81751
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-465 <TET>
A:Cross-references: GB:AE002269; GB:AE002160; NID:g7190041; PIDN:AAF38896.1; PID:g719
A:Experimental source: strain Nigg (MoPn)
C:Genetics:
A:Gene: TC0002
C:Superfamily: Vibrio alginolyticus sodium-translocating NADH dehydrogenase (ubiquino
C:Keywords: oxidoreductase
```

```
alignment_scores:
Quality: 453.00 Length: 464
Ratio: 1.520 Gaps: 14
Percent Similarity: 64.224 Percent Identity: 28.448
```

```
alignment_block:
US-09-303-518D-131 x C81751 ..
Align seg 1/1 to: C81751 from: 1 to: 465
```



4 ATTTAAATCAAAAAAGGTCCTAAATCTGCCCATCGCGGGCAGACCG...GAT

3 IleThrValAsnArgGlyLeuAspLeuSerLeuGlnGlySerProLysG1 19  
51 GCAAGTCATTATGACGGCCGGCCATTACCGAAGTCGGTTCGTTGGCG 100  
19 uSerGlyPheTyrAsn.....LysIleAsp 28  
101 AGAATATGTCGC.....ATCGGCCCC.....TCG 126  
28 roGluPheValSerIleAspLeuArgProPheGlnProLeuSerLeuLys 44  
127 ATGAATCAAGGAAGTGAAGCGTCAAAAGGCGCAAGTGTGTTGA 176  
45 LeuLysValGlnGlnGlyAspAlaValCysSerGlyAlaProIleAlaG1 61  
177 AGCAAAAAGATCCGGCGGTAGTATTACTGCGCGCGGTTCAGGCAAAA 226  
61 uTyrLysHisPheProAsnThrTyrIleThrSerHisValSerGlyValV 78  
227 TCAGCGCTATTCACCTGCGGAAAGCGTACTTCCAGTCAGTCGTGATT 276  
78 alThrAlaIleArgArgGlyAsnLysArgSerLeuLeuAspValIleIle 94  
277 ...CGCGTTGAAGGCAACGACGAAATCGAGTTCGAACGCTACCTGA 323  
95 LysLysThrProGlyProThrSerThrGluTyr...ThrTyrAspLeuG1 110  
324 AGCGCTGGCAAAATTCAGCAGCAAAAGTGCAGCGCAACCTGATTCAT 373  
110 nThrLeuSerArgSerAspLeuSerGluIlePheLys.....GluA 124  
374 CAGCGTTATGAGTTCGCTCGCACCGCTCGCTTCAGCAAAATCCCTGCC 423  
124 sGlyLeuPheAlaLeuIleGlyGlnArgProPheAsp...IleProAla 139  
424 GTAGATCCGAG...CGTTTCGCATCTTCGTCAATGCGAGTGACACCAA 470  
140 IleProThrGlnThrProArgAspValPheIleAsnLeuAlaAspAsnAr 156  
471 TCAGCTGCTGCGACCTACGCTACATC.....A 502  
156 gProPheThrProSerProGluLysHisLeuAlaLeuPheSerSerArg 173  
503 AAGAAGCCGCGAGACTTCAACCGCGCTGTTGTTGTTATGACCGCGCTG 552  
173 luGluGlyPheTyrValPheValValGlyValArgAlaIleAlaLysLeu 189  
553 ACCGACGTAAATCCATGTGTGTAAAGCAGCAGCGCCACAGCTGCGTC 602  
190 PheGlyLeuArgProHisIleValPheArgAspArgLeuThrLeuProTh 206  
603 TGAATATGCTGCCAATATC...GAACACATGAATTTGGCGCGCGCATC 649  
206 rGlnGluLeuLysThrIleAlaHisLeuHisThrValSerGlyProPheP 223  
650 CTGCGCGCTGAGTGGCAGCAGCATTCATTCATCGAGCCAGTCGCGCGG 699  
223 roSerGlySerProSerIleHisIleHisSerValAlaProIleThrAsn 239  
700 AATAAA...ACCGTGTGACCATCAATATCAAGACGTGATGCTATCGG 746  
240 GlnLysGluValValPheThrLeuSerPheGlnAspValLeuThrIleG1 256  
747 ACGTTTGTTCGTAACAGCGCTCTGAATACCGAGCGGTGTTGCCCTGG 796  
256 yHisLeuPheLeuLysGlyArgIleLeuHisGluGlnValThrAlaLeuA 273  
797 GCGGC.....CTGCAAGTCAACAAACCGCGCTCTTCGCTACCGTTTG 840  
273 laGlyThrAlaLeuLysSerSerLeuArgArgTyrValIleThrLys 289  
841 GGTGCGAGGTGCTCACTACCTACCGCGCGCAATGTTGTCAGCGGACAA 890  
290 GlyAlaSerPheSerSerLeuAsnLeuAsnAspIleSerAspAsnAs 306

891 CCGCGTATTTCCGTTCCGTTATTAACGFTGCGATTGTCACAGGCGCGC 940  
306 pThrLeuIleSerGlyAspProLeuThrArgLeuCysLysLysGluG 323  
941 ATGAT...TATTTGGAGCGTACACCAATCAGATTTCCGTTATCGAAGAA 987  
323 luGluProPheLeuGlyPheArgAspHisSerIleSerValLeuHisAsn 339  
988 GCGCGCAGCAAGAGCTGTTCCGCTGGTTCGCGCGCGAGCGGACAAATA 1037  
340 ProThrLysArgGluLeuPheSerPheLeuArgIleGlyPheAsnLysPr 356  
1038 CTCCATCAGCGCACCTCTCGGCCATTTCTTAAACAACTCTTCA 1087  
356 oThrPheThrLysThrThrLeuSerGlyPhePheLysLysArg...T 372  
1088 AGTTACG.....ACAGCGTCAACGCGCGCGCGCGCATCGTA 1128  
372 hrTyrThrAsnProAspThrAsnLeuHisGlyGluThrArgProIleIle 388  
1129 CGATCGCGCACTTATGACGCGTAATCGCGTGGACATCCTGCTACCTT 1178  
389 AspThrAspIleTyrAspLysValMetProMetArgIleProValValPr 405  
1179 GCTTTTGGCGGATTTAATCGTCGCGGATACCGACAGCGCGCGCTTGG 1228  
405 OleuIleLysAlaValIleThrLysAsnPheAspLeuAlaAsnGluLeuG 422  
1229 GTTGTTCGATTTGACGAAGAAGACCTCGCTTTGTCAGCGTTCGCTGC 1278  
422 lPheLeuGluValCysGlyGluAspPheAlaLeuProThrLeuIleAsp 438  
1279 CGGCGCAATACGAATACGCGCGCTGTCGCAAAAGTGTGGAACCAT 1328  
439 ProSerLysThrGluMetLeuThrIleValLysGluSerLeuIleGluTy 455  
1329 TGAGAAGGAA 1338  
455 rAlaLysGlu 458  
seq\_name: pir2:E72040

seq\_documentation\_block:  
probable sodium-translocating NADH dehydrogenase (ubiquinone) (EC 1.6.5.-) alpha chain  
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 16-Feb-2001  
C:Accession: E72040; G81623  
R:Kelman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, R.; Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:99206606  
A:Accession: E72040  
A:Molecule type: DNA  
A:Residues: 1-467 <ARN>  
A:Cross-references: GB:AE001656; GB:AE001363; NID:g4377047; PIDN:AAD18882.1; PID:g437  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39  
A:Reference number: A81500; MUID:20150255  
A:Accession: G81623  
A:Molecule type: DNA  
A:Residues: 1-467 <REA>  
A:Cross-references: GB:AE002164; GB:AE002161; NID:g7188939; PIDN:AAF37899.1; PID:g718  
A:Experimental source: strain AR39, HL cells  
C:Genetics:  
A:Gene: nqrA; CP0002  
C:Superfamily: Vibrio alginolyticus sodium-translocating NADH dehydrogenase (ubiquino  
C:Keywords: oxidoreductase



C;Genetics:  
A:Gene: VC1015

C/Genes: VC1015  
A:Gene: VC1015



A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-673 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD01910.1; PID:g16502752; GSPDB:GN00176  
C:Genetics:  
A:Gene: STV1665  
C:Superfamily: unassigned ferredoxin 2[4Fe-4S]-related proteins; ferredoxin 2[4Fe-4S] h  
alignment\_scores:  
Quality: 151.50 Length: 449  
Ratio: 0.728 Gaps: 17  
Percent Similarity: 46.325 Percent Identity: 22.049  
alignment\_block:  
US-09-303-518D-131 x AE0692 ..  
Align seg 1/1 to: AE0692 from: 1 to: 673  
100 GAAGAATATGTCGGCATCGCCCTCGATGAAATCAAGAAAGGTGAAGC 149  
.....  
48 LysGlnHisIleGlyAlaGluGlyGluLeuCysValSerValGlyAsp 64  
.....  
150 CGTCAAAAGGCAAGTGTCTTTGAAGACAAAGAAAGATCCGGCGTAG 199  
||| .....  
64 gValLeuArgGlyGlnAlaLeuThrArgGlyArgGlyArgMetLeuProV 81  
.....  
200 TATTACTGCGCGGCTTCAGGCAAAATCGCCCTATT ..... 237  
|| .....  
81 alHis...AlaProThrSerGlyThrValIleAlaIleAlaProHisSer 96  
.....  
238 .....CACGTCGGCGAAAGCGCTACTTCAGTCAGTCGTCGATTCGCGT 281  
|| .....  
97 ThrAlaHisProSerAlaLeuAlaGluLeu...SerValIleIleAsp 112  
.....  
282 TGAAGGCAACGACGAA .....ATCGAGTTCGAAC 310  
.....  
112 aAspGlyGluAspArgTrpIleGluArgGluGlyTrpSerAspTyrArg 129  
.....  
311 GCTACGTACTGAGCGCTGCAAAATGAGCAGCGAAAGTGCAGCGC 360  
||| .....  
129 laHisSerArgGluAlaLeuIleGluArgIleHisGlnTyrGlyValAla 145  
.....  
361 AACCTGATTCAATCAGCTTATGAGTCGCTTCGACCGCTCGACCGCTCGTAC 410  
||| .....  
146 GlyLeuGlyGlyAlaGlyPheProThrGlyVal ..... 156  
.....  
411 CAATCCCTGCGCTAGATCCGAGCGGTTCGCCATCTTCGTCATCGCA 460  
||| .....  
157 .LysLeuGlnGlyGlyAspAspLysIleThrThrLeuIleIleAsnAla 173  
.....  
461 TGGACACCAATCGCTGCTGCGGACCCCTACGGTCATCATCAAGAAAGCC 510  
||| .....  
173 laGluCysGluProTyrIleThrAlaAspArgLeuMetGlnAspCys 189  
.....  
511 GCCGAGACTCAACACGCGCTTTGGTATTGAGCGCGCTGACCGAA... 558  
||| .....  
190 AlaAlaGlnIleValGluGlyIleArgIleLeuAlaHisIleLeuGlnPr 206  
.....  
559 .....CGTAAATCCATG 571  
||| .....  
206 oArgGluValLeuIleGlyIleGluAspAsnLysProGlnAlaIleSerM 223  
.....  
572 TGTCTAAGCAGCAGCGCGCTGCTGAAATGCTGCAATATC 621  
||| .....  
223 etLeuArgAlaValLeuAlaAsp .....AlaHisAspIle 234  
.....  
622 GAAACACATGAATTTGGCGCGCGCATCTCCGCGGC ..... 657  
||| .....  
235 SerLeuArgValIleProThrLysTyrProSerGlyGlyAlaLysGlnLe 251  
.....  
658 .....TTGAGTCGCGCAGCATTCATTTCATCGAGCCAGTCGCGC 697  
||| .....  
251 uThrGlnIleLeuThrGlyLysGlnVal .....ProHisGly 264  
.....  
698 CGATAAAGCGGTGTGACCATCAATTTATCAAGAGGTG .....ATT 738  
||| .....  
264 lYArgSerSerAspIleGlyValLeuMetGlnAsnValGlyThrAlaTyr 280  
.....  
739 GCTATCGAGCTTTGTTTCGTAACAGCGCGTCTGAATACCGAGCGGTGT 788  
||| .....  
281 AlaValLysArgAlaValValAspGlyGluProIleThrGluArgValVa 297  
.....  
789 TGCCTTGGCGGCTGCAAGTCAACAAACCGCGCTCTTCGTCACCGTTT 838  
||| .....  
297 lThrLeuThrGlyGluAlaIleSerArgProGlyAsnValrPalargL 314  
.....  
839 TGGGTGCGAAGGTGTCTCAACTT .....ACCGCGCGCGAATTGGTTGAC 882  
||| .....  
314 euGlyThrProValArgHisLeuLeuAsnAspAlaGlyPheCysProser 330  
.....  
883 CGGACAAACCGGTGATTTCGGTTCGGTATTGAACGTCGTCGATTGCACA 932  
||| .....  
331 AlaAspGlnMetValIleMetGlyGlyProLeuMetGlyPheThrLeuPr 347  
.....  
933 AGCGCGCATGATTATTGGGACGCTACCAACAATCAGATT .....T 973  
||| .....  
347 oTrpLeuAspValProValValLysIleThrAsnCysLeuLeuAlaProS 364  
.....  
974 CCGTTATCGAAAGAGCGCGCAAGAGCTGTTCGGCTGGGTGGCGCG 1023  
||| .....  
364 erValThrGluMetGly .....AlaPro 371  
.....  
1024 CAGCGGACAAATCTCCATCCGCGCACCACTCTCGGCCATTTCCTAAA 1073  
||| .....  
372 GlnGluGluLysSerCysIleArgCysSerAlaCysAlaAsp ..... 385  
.....  
1074 AAACAAACTCTTCAAGTTTCAGCAGCGCTCAAGCGCGCGCGCGCCCA 1123  
.....  
385 ..... 385  
1124 TGGTACCGATCGGCACCTTATGAGCGGTAAATGCGGTGGACATCTGCCT 1173  
||| .....  
386 .....AlaCysProAlaAspLeuPro 393  
.....  
1174 ACCTGCTTTTCGGGATTAATCTCGCGGATACCGACAGCGCGCAGGC 1223  
||| .....  
394 GlnGlnLeuTyrTrpPheSerLysGlyGlnGlnHisaspLysAlaThrAl 410  
.....  
1224 T .....TTGGGTGCTTGGATTTGGACGAAGACCTCGCTTGT 1264  
||| .....CysGlyAlaC 422  
410 ahHisIleAlaAspCysIleGlu .....  
1265 GCAGCTTCGTCGCGCGCAATACGAATACGCGCGCTGTCGCGC 1311  
||| .....ProLeuValGln 434  
422 ysAlaTrpValCysProSerAsnIle .....  
seq\_name: pir2:S59310  
seq\_documentation\_block:  
probable membrane protein YMR317w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein YMR924.09  
C:Species: Saccharomyces cerevisiae  
C:Date: 29-Nov-1995 #sequence\_revision 23-Feb-1996 #text\_change 05-Nov-1999  
C:Accession: S59310  
R:Churchev, C.M.  
submitted to the EMBL Data Library, September 1995  
A:Reference number: S59302  
A:Accession: S59310  
A:Molecule type: DNA  
A:Residues: 1-1104 <CHU>  
A:Cross-references: EMBL:Z54141; GSPDB:GN00013; MIPS:YMR317w  
A:Experimental source: strain AB972  
C:Genetics:  
A:Gene: MIPS:YMR317w  
A:Map position: 13R

[illegible]

seq documentation block:

seq documentation block:

hypothetical protein KIAA0324 - human (fragment)

C;Species: Homo sapiens (man)  
Information Process: KIRK0324  
human (fragment)

C;Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999

C;Accession: T02345

R; Ricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.

re, J.; White, S.; Ueng, S.; Tatum, O.; Campbell,

submitted to the EMBL Data Library, March 1998

A;Description: Sequencing of human chromosome 16

A;Reference number: Z14664

A;Accession: T02345

A; Status: preliminary; translated from GB/EMBL/D

A;Molecule type: DNA  
A.Residues: 1-1701

A; residues: 1-1791 <RIC>  
A: Cross-references: EMBL: AC004403. NID: 53005548.

A/Cross-References: EMBL:AC004493; NID:g2996648;  
C:Genetics:



A:Map position: 16  
A:Introns: 1610/2; 1706/2  
A:Note: KIA0324

alignment\_scores:  
Quality: 141.00 Length: 446  
Ratio: 0.701 Gaps: 17  
Percent Similarity: 45.067 Percent Identity: 25.356

alignment\_block:

US-09-303-518D-131 x T02345 ..

Align seg 1/1 to: T02345 from: 1 to: 1791

```

32 CCATCCGGGACACCGGAGCAAGTCATTTATGACGGCCCGCCATTACC 81
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
753 ProProArgHisArgArgSerProSerValSerProGluProAlaG1 769
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
82 GAAGTCGGCTGCTGGCGAAGATATGTCGGCATGCGCCCTCGATGAA 131
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
769 uLysSerArgSerArgArgArgSerAlaSerProArgThrL 786
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
132 AATCAAGGAAGTGTAAGCGCTCAAAAAGGCCAAGTCTGTTTGAAGACA 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
786 yThr .....Thr 788
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
182 AAAAGATCCGGCGTAGTATTACTGCGCGGCTTCAGGCAAAATCGCC 231
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
789 SerArg, ArgGlyArg ..... 793
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
232 GCTATTCACGTCGGCGAAGACGCGTACTTCAGTCAGTCGTGATTCGCT 281
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
794 ....SerProSerProLysProArg....GlyLeuGlnArgSer....Arg 806
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
282 TGAAGCAACGACGAATCGAGTTCGAACGCTACGTACCTGAAGCGCTGG 331
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
807 SerArgSerArgGluLysThrArgThrArgArg ..... 819
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
332 CAANAATGACGAGAAAGTGGCGGCAACCTGATTCAATCAGCGCTTA 381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
820 .....ArgAspArgSerGlySerGlnSerThrSerArgArg.... 832
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
382 TGGACTCGCTTGGCACCGCTCGCTTCAGCAAAATCCCTCCGCTAGATGC 431
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
833 .....ArgGlnArgSerArgSerArgSer 840
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
432 CGAGCGCTTCGCATCTTCGTAATCGGATGGA...CACCAATCCGGTGG 478
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
841 ArgValThrArgArgArgGlyGlySerGlyTyrHisSerArgSerPr 857
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
479 CTGCGGACCTTACGTCATCATCAAGACGCGCGA.....AGAC 519
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
857 oAlaArgGlnGluSerSerArgThrSerSerArgArgArgGlyArgS 874
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
520 TTCAACGCGGCGCTGTGGTATTGAGCGCGCTGACCGAAGCTAAATCCA 569
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
874 erArgThrProThrSerArgLysArgSerArgSerArgThrSerPro 890
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
570 TGTGTGTAAACGACGCGCGACGCTGCGCTGCTGAAATGCTCCCAATA 619
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
891 AlaProTrpLysArgSerArgSerArgAla..... 900
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
620 TCGAAACACATGAATTTGGCGCGCGGATCTGCGCGCTTGAGTGGCAGC 669
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
901 .....SerProAlaThr.....HisA 906
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
670 CACATTCATTTATCGCA.....GCCAGTCGCGCGGATATAAAC 707
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
906 rgArgSerArgSerArgThrProLeuLeuSerArgArgSerArgSer 922
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
708 CGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGACGCTTGTTCG 757
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

923 ArgThrSerProValSerArgArgSerArgSerArgThrSerValTh 939
758 TACAGCGCGCTCTGAATACGAGCGCGTGGTTCCTTGGCGCGCTGCAA 807
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
939 rArgArgArgSerArgArgAlaSerProValSerArgArgArgSera 956
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
808 GTCAACAACCGCGCTCTTGGCTACGCTTGGTGGTGGTGGTGGTCTCA 857
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
956 rgSerArgThrProProValThrArgArgSerArgSerArgThrPro 972
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
858 ACTTACCGCGCGCAATTGTTGACGCGGACAAACCGCGTATTCGCGTT 907
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
973 ThrThrArgArgSerArgSerArgSerArgThrPro..... 984
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
908 CGGTATTGAACGGTGCATTCACAGCGCGCATGATTATTTGGGACGC 957
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
985 .....ValThrArgArgArgSerArgSerArgThrP 995
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
958 TACCACATCAGATTTCGGTATTCGAAGAGCGCGCAGCAAGAGCTGTT 1007
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
995 roProValThrArgArgSerArgSerArgThrSerProIleThrArg 1011
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1008 CGGCTGGGTTCCGCGCGCGGCAAAATCTCCATCAGCGCACCACTC 1057
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1012 ArgArgSerArgSerArgThrSerProValThrArgArgSerArgSe 1028
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1058 TCGG.....CCATTTCCTAAAAAACAACCTCTCAAGTTCAACGACGC 1101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1028 rArgThrSerProValThrArgArgSerArgSerArgThrSerProV 1045
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1102 GTCAACGCGCGCGCGCGCGCTGTCACCGATCGGCTATGAGCGCT 1151
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1045 alThrArgArgSerArgSerArgThrProProAlaIleArgArgArg 1061
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1152 AATCGCTTGGACATCCCTACCTCTTTCGCGGATTTAATCTCGTCG 1201
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1062 SerArgSerArgThrProLeu.....LeuProArgLg 1072
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1202 GCGATACCGACGCGCGCGCTTGGTTCGTTGGATTCGACGCA...A 1248
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1072 sargSerArgSerArgSerProLeuAlaIleArgArgSerArgSera 1089
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1249 GAAGACCTCGCTTGTGCGAG.....CTTCGCTCGCGCGGG 1283
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1089 rgThrProArgThrAlaArgGlyLysArgSerLeuThrArgSerPro 1105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1284 CAAATACGAATACGCGCGCTTTCGCGCAAAAGTCT 1319
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1106 AlAlaIleArgArgArgSerAlaSerGlySerSer 1117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

seq\_name: pir2:E64136

seq\_documentation\_block:

infC protein homolog - Haemophilus influenzae (strain Rd KW20)  
C:Species: Haemophilus influenzae  
C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 17-Mar-2000  
C:Accession: E64136  
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage  
Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman  
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter  
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630  
A:Accession: E64136  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-819 <TIGR>  
A:Cross-references: GB:U32841; GB:U4203; NID:G1574529; PIDN:AA023331.1; PID:G1574537  
C:Superfamily: unassigned ferredoxin 2[4Fe-4S]-related proteins; ferredoxin 2[4Fe-4S]  
F:371-435/domain: ferredoxin 2[4Fe-4S] homology <FER>

```

Length: 465
Quality: 135.50
Ratio: 0.622
Gaps: 15
Percent Similarity: 46.882    Percent Identity: 19.570

alignment_block:
US-09-303-SIBD-131 x E64136 ..

Align seg 1/1 to: E64136 from: 1 to: 819

   70 CCGGCCATTACCGAAAGTCGGTTGCTTGGCGAAGAATATGTCCGCATGCG 119
      |||:::
      |||:::
   38 ProLeuglyThrAspPheTyrlleProLeuLysGlnHisLeuGlyIleThrH 54
      :::::
   120 CCCTCGATAAANAACAAGAGGTGAAGCCTCAAAAAGGCCAACGTCC 169
      ::::: |||||:::: |
   54 rGIyASnLeuLeuIIeLysGlucLyAspYrValLeuLysGlyGlnAla 71
      |||:::
   170 TGTTTTGAAAGACAAAAAGATCCGGGCGTAGTATTACTCGCGCGCGCTTCA 219
      |||:::
   71 euThrILysGLyASPGLyLeuArgMetLeuProValHisAlaProThrSer 87
      |||:::
   220 GGC AAAATCGCCGCTATTACC CGTGGCGAAAAGCCGCTACTTTCAGTCACT 269
      ||| |||
   88 GlyThrIle.....LysSerILELysProTYrVa 97
      |||:::
   270 CGTGATTGCGGTTGAAGGCAACGACGAANATCAG ..... 303
      |||:::
   97 lAlaThrHisProSerGlyLeuASpGUlUProThrIleHisLeuGlnAla 114
      |||:::
   304 .....TYCGAACGCTAGTACCTGGAAGCGCTGCCAAAA 336
      |||:::
   114 spGLyLeuASPGLNTrpIleGluArgSnProILEAspPheSerThr 130
      :::::
   337 TTGAGCAGCAAAGATCGCGCGCAACCTGATTCAATCAGGCTTATNGAC 386
      |||||||:::::
   131 LeuSerSergLUgInLeuIleHisLysILEtyrGlnAlaGlyLE...Al 146
      |||:::
   387 TGCGCTTCGACCCGCTCGCGTTC.....AGCAAAATCCCTCGCGGTAG 427
      |||:::
   146 aGLyLeuGlyALyAlavalPheProThrAlaAlaySIleGlnSerAlag 163
      |||:::
   428 ATCCGCGCCGTTCCGCATCTTCGTCAATGGATGGACACCAATCCCGTG 477
      |||:::
   163 luGLINlysValLysLeuLeuILElleISanGlyAlaGLUCysGLUProTYr 179
      |||:::
   478 GCTGCGCACCCCTACGCTCATCATCAAGACGCCGCGAAGACTTCAACAG 527
      |||:::
   180 lleThrCySAspASPARGLeuMetargLuargAlaspgLIullelleLY 196
      |||:::
   528 CGGCTGTGTGTTGACGCGCGTACCGAAGACGTAAATCCATCGTGTGA 577
      |||:::
   196 sgLYleArgIleLeuArGYrIleLeuHisProGLU...LysValVall 212
      |||:::
   578 AAGCAGCAGGCGCAGCGTCCGCT..... 603
      |||:::
   212 leAlalleLyuAspnLysProglualalleSerAlalleArgASNala 228
      |||:::
   604 ...GAAATGCTGCCAATATCAACACATGAATTTGGCGCGCGCGCATCC 650
      |||:::
   229 LeuGINglyAlaAsnAPileSerILEArgvallleProthrLysTyRPr 245
      |||:::
   651 TCGCGGCTTGAGTGGCAGCGCATTCATTCATTCAG.....C 688
      |||:::
   245 oSERGlyAlathrLysGlnLeuILLEtyrLeuLeuthrglyleGLUvalP 262
      |||:::
   689 CAGTCGGCGGCAATAAACCGTGTGGACCATCAATATCAAGACGTG... 735
      |||:::
   262 roSERGlyGuLuarSerSerSertIleGlyValLeuMetGlnAnsnValGLy 278
      |||:::
   736 .....ATTGCTATCGCAAGCTTTGTCGTACAGCGCGCTGTAATACCCA 779
      |||:::
   279 ThrMetPhelAlleLysArgAlaIllealeASnAspGUleProLeuILEGI 295

```

alignment scores:

819 GCGCTCTTGGT.....ACCGTTTGGTGGCAAGTGTCTCAAC 859  
273 yValValGlnAsnValGlyThrCysValAlaValLysGluAlaValV 290  
860 TTACCGCGCGGAATTTGGTTGACGGGACACCGCGGTGATTTCCGGTTCG 909  
290 aAspGlyLysProLeuValGluArgGlyMetThrValSerGlyAspAla 306  
910 GTATTGAAC.....GGTGGATTGCACAGG 935  
307 ValLysAsnGlnLysAsnLeuLeuValArgIleGlyThrProValLys 323  
936 CGCGCATGATTATTGGGACGCTTACCAATCAGATTTCGGTTATCGAAG 985  
323 pValIleAspTyrCysGly.....IleAspG 333  
986 AAGCGCGCAGCAAGAGCTTTCGGTGGTTCGCGCGGCGGACCGGACAAA 1035  
333 luAsnThrGluArgValIleLeuGly.....GlyProMetMetGly... 346  
1036 TACTCCATCAGCGCACCACTCTCGGCGCATTTCTTAAACAACTCTT 1085  
347 lIleSerIleThrAsn.....LeuAspIleProValMe 357  
1086 CAAGTTCACGACGCGCTCAACGCG.....GGCG 1114  
357 tLysGlyThrSerGlyIleThrAlaPheLeuProLysLysSerArgProG 374  
1115 ACCGCGCATGGTACCGACGCTTATGAGCGCGTAAATCGCGTTGGAC 1164  
374 lInysProCysIleArgCysSerGluCysValGlnValCysProMetAsn 390  
1165 ATCTGCTACCTTCTTTCGCGGATTTAATCGCGGCGATACCGACAG 1214  
391 LeuGlnProTyrLeuLeuTyrLeuSerThrLysArgLysTyrAspG 407  
1215 CGCGCGCTTGGTGGTGGTGAATGGAGAGAGACCTCGCTTTGT 1264  
407 uAlaValGluAsnGlyLeuMetAspCysIleGlu.....CysGlySerC 422  
1265 CGAGTTCGTCGCGCGGCAAAATACGAATAC 1296  
422 yThrTyrThrCysProSerLysIleGluHis 432

seq\_name: pir2:D83208  
seq\_documentation\_block:  
probable ferredoxin PA3491 [imported] - pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: D83208  
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
.; Lory, S.; Olson, M.V.  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
A:Reference number: A82950; MUID:20437337  
A:Accession: D83208  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-774 <STO>  
A:Cross-references: GB:AE004770; GB:AE00491; NID:99949633; PIDN:AG06879.1; GSPDB:GN  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA3491

alignment\_scores:  
Quality: 132.00 Length: 492  
Ratio: 0.600 Gaps: 24  
Percent Similarity: 44.715 Percent Identity: 20.935

alignment\_block:

Align seg 1/1 to: E72398 from: 1 to: 451  
49 GAGCAAGTCATTATGACGCGCGCCGCTTACCGAAGTCGCTGCTTGG 98  
35 AspLysProIleGluArgAlaProLeuProGlnLysValPheValPheLe 51  
99 CGAAGATATCTCGCATCGCGCCCTCATGAAATCAAGGAGTGAAG 148  
51 uSerAsnHisAlaGlyAsnProAlaLysProValValSerProGlyAspG 68  
149 CCGTCAAAAAGCGCAAGTCTGTTTGAAGACAAAAGAAATCCGGCGCTA 198  
68 luValLysThrGlnValIleGlyGluProGluGlyPheIleSerAla 84  
199 GTATTACTCGCGCTTACGCAAAATCGCGCTATTCACCGTGGCGA 248  
85 TyrLeuHisSerProValThrGlyArgValLeuGluIle..... 97  
249 AAGCGCGTACTTACGTAGTCTGCTG.....ATTGCGG 280  
98 .LysLysIleLeuHisProIleLeuGlyLysProIleGluAlaIleValI 114  
281 TTGAAGGCAACGACGAAATCGAGTTCGACGCTAGCTGACGCGCTG 330  
114 leGluArgThrSerAspGluTrpValHisIleGluThrGlyAspPhe 130  
331 GCAAAATGAGCAGCAAAAGTGGCCGACCTGATTCATCA 375  
131 GluArgMetSerLysGluLeuLeuGluIleLysLysAlaGlyI 147  
376 ....GCTTATGAGTCTGCGCTTACGCGCTGCTTCAAGAAATCCCTG 421  
147 eValGlyLeuGlyGlyAlaMetPheProThrHisValLysLeuSerProp 164  
422 CGTAGATCGGAGCGTTCGCGCTTTCGTAAGTCGATGACGACCAAT 471  
164 roProGluLysValAspThrLeuIleValAsnGlyAlaGluCysGlu 180  
472 CCCTGCTGCGGACCTTACGCTATCAAAAGAGCGCGGACGACT 521  
181 ProValLeuThrIleAspHisArgLeuMetLeuGluArgAlaGluAsp 197  
522 CAACGCGGCTGCTGTTGATTTAGCGCGCTGACCGCAACGTAATCCATG 571  
197 eLeuGlnGlyLeuLeuIleMetLysVal.....LeuGlyV 210  
572 TGTGTAAGCA...CGAGCGCGACAGTGGCTCTGAAATGCTGCAAT 618  
210 alGlnLysAlaValValGlyValGluSerAsnLysMetAspAlaTyrHis 226  
619 ATCGAAACACATGAATTTGGCGCGCGCATCTGCGGCTTGTAGTGGCAC 668  
227 AsnLeuLysLysValPheLysGly...TyrProValAspValAlaLeuLe 242  
669 GCACATTCATTCATCGACGCTGCGCGCGGATTAACCGCTGTGACCA 718  
242 uArgThrLysTyr.....ProGlnGlyAlaGluLys..... 252  
719 TCATTATCAAGAGGTGATTGCTATCGGACGCTTTGTTGTAACAGCGCGT 768  
253 .....GlnLeuIleValAlaIle..... 258  
769 CTGAATACCGGCGCGGTGCTTGGCGGCGCTGCAAGTCAACAAC 818  
259 .....ThrGlyArgMetValProArgGlyGlyLeuProMetAspValG 273

US-09-303-518D-131 x D83208

Align seg 1/1 to: D83208 from: 1 to: 774

28 CTGCCCATCGCGGCGAGAGCGGAGAGTATTTATGAGCGCGCGCCGAT 77  
|||||  
23 LeuProIleGlnGlnAlaProLeuAlaGlnArgTyrIleValPro. .... 37  
78 TACCGAAGTCGCGTTCCTTGGCGGAAGATATATGTCGCC ..... ATGCGCC 121  
|||||  
38 ..... LeuGly. .... GlnHisIleGlyAlaProAlaAArgp 48  
122 CTCTCATGAAATCAAGAGTGAAGCGTCAAGCGTCAAAAAAGCGCAAGTGTG 171  
|| .....  
48 rocysValGluVal ..... GlyGlnAlaValLeuLysGlyGlnThrIle 62  
172 TTTGAAGCAAAAAGATCCCGCGCTAGTATTACTGCGCGGCTTCAGG 221  
|| .....  
63 AlaLeuProAspGlyThrValSerAlaAlaLeuHisAlaProThrSerGI 79  
222 CAAATCGCGCGTATT ..... CACCGTGGCGAAGAAAGCGCG 256  
| .....  
79 yThrValValAlaIleGlyAlaHisProTyrProHisAlaSerGlyLeup 96  
257 TACTTCAGTCAGTCGTGATTGCCGTTGAAGCAAGCAGCAA ..... 297  
|||||  
96 roAlaProAlaIleValIleAlaSerAspGlyLeuGluArgTyrThrGlu 112  
298 ..... ATCGAGTTCGAAGCGTACGTACTGAAGCGCTGGCCAA 335  
|||||  
113 LeuHisProCysProAspPheArgAlaGluSerProLeuAlaLeuLeu.. 128  
336 ATTGAGCAGCAAAAAGTGCGCCG ..... AACCTGATTCAAT 373  
|||||  
129 ..... GluArgIleArgAlaAlaGlyIleGlyLeuGlyAla 142  
374 CAGGCTTATGACATCGG ..... CTTCGCCACCGTCTGCTTCAAGAAATC 417  
|||||  
142 laGlyPheProThrAlaAlaLysLeuAlaAlaArgProAlaGluLysIle 158  
418 CTGCCGTAGATGCCGAGCGCTTCGCATCTTCGTCAATGATGGATGGACAC 467  
|||||  
159 ..... HisThrLeuValValAsnGlyAlaGluCys 168  
468 CAATCCGCTGCTCGGACCGTACGGTATCATCAAAAGAGCGCGCGAAG 517  
|||||  
168 sGluProTyrIleSerAlaAspLeuLeuMetArgGluArgAlaThrG 185  
518 ACTTCAAAAGCGCGCTGTGTTATGAGCGCGCTGACCGACGAGTAAATC 567  
|| .....  
185 InValLeuGlyGlyLeuAspIleLeuValGlnIle ..... 196  
568 CANTGTGTAAAGCAGCAGCGGACGCGTGCCTGCTGAAATGCTGCCAA 617  
|||||  
197 ... Leucys ..... ProGluGluValLeuValGI 205  
618 TATCGAACAACATGAA ..... TTGCGCGCGCG ..... 645  
|||||  
205 yIleGluAspAspLysProGluAlaIleAlaAlaLeuGlyAlaAlaLeuG 222  
646 ..... CATCTCGCGCGC 657  
222 LYGluArgProTyrArgIleValAlaLeuProThrArgTyrProSerGly 238  
658 ..... TTGAGTGGCAGCAGCATTCATTTTCAT 683  
239 GlyIuArgGlnLeuIleGlnLeuLeuThrGlyArgGluValPro..Al 254  
684 CGAGCCAGTCGCGCGCAATAAACCCTGTGGACCATCAATTATCAAGACG 733  
|||||  
254 aspGlyLeuProAlaAspIleGlyIleLeuCysGlnAsnValGlyThrL 271  
734 TGATTGCTATCGACGCTTTGTTTCGTAACAGCGCGTCTGAATACCGAGCGC 783

271 euAlaAlaValHisAspAlaValValLeuGlyArgProLeuIleSerArg 287  
|||  
784 GTGGTTCCTTGGCGCGCTGCAAGTCAACAAACCGCGCTCTTGGCGTAC 833  
|||||  
288 IleThrThrLeuAlaGlyGlyAlaLeuGluArgProMetAsnValGluAl 304  
834 CTTTTGGTCCGAAGGTGTCTCAACTTACCGCG ..... 867  
|||||  
304 aLeuIleGlyThrProValHisGluLeuLeuAlaPheAlaGlyLeuAlaG 321  
868 ..GCGGAATTGGTACGCGCAGACCGCTGATTTCCTCGGTTCGGTATTG 915  
|||||  
321 luGlyArgLeu ..... GluArgValLeuMetGlyGly ..... 331  
916 AACGGTCGATTGCACAAGCGCGCATATTATTGGGACGCTACCACAA 965  
331 ..... 331  
966 TCAGATTTCGTTATCGAAGAAGCGCGCAGCAAGAGCTGTTCCGCTGGG 1015  
|||||  
332 ..... ProMetMetGlyPheA 337  
1016 TTGCGCGCAGCGGACAAATACTCCATCACGCGCACCACCTCTCGGCCAT 1065  
|| .....  
337 laLeuProAsp ..... 340  
1066 TTCTTAAAAAACAACCTTCAAGTTACGACAGACCGCTCAACGCGCGCA 1115  
|||||  
341 ...LeuSerValProLeuIleLysThrCysAsnCysLeuLeuAlaGlyAs 356  
1116 CCGCGCATGCTACCG ..... ATCG 1135  
356 pAlaThrGluLeuProGluProValProAlaMetProCysIleArgCysG 373  
1136 GCATTATGAGCGCGTAATGCGCTTGGACATCTGCTACCTTGTGCTTTG 1185  
|||||  
373 LyAspCysAlaGlnValCysProValSerLeuLeuProGln...GlnLeu 388  
1186 CCGGATTATTCGTCGGGATACCGACAGCGCGAGGCTTTGGTGTGCTT 1235  
|| .....  
389 HisPheAlaLeuGlyAsp...GluHisGluGlnLeuLeuAlaHisAs 404  
1236 GGAATTGGACGAAGACCTCGCTTTGTGACAGCTTCGCTGCGCGGCA 1285  
|||||  
404 nLeuPheAspCysIleGluCysGlyAlaCysAlaTyrValCysProSerS 421  
1286 AATACGAATACGCGCGCTGTTGCGC 1311  
|| .....  
421 erIle ..... ProLeuValGln 426

seq\_name: pir2:A84943

seq\_documentation\_block:

membrane protein [imported] - Buchnera sp. (strain APS)

C:Species: Buchnera sp.

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001

C:Accession: A84943

R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.

NATURE 407, 81-86, 2000

A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp

A:Reference number: A84930; MUID:20445173

A:Accession: A84943

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-473 &lt;STO&gt;

A:Cross-references: GB:AP000398; GSPDB:GN00144

A:Experimental source: strain APS

C:Genetics:

A:Gene: rnfC

alignment\_scores:

Quality: 131.00 Length: 440  
 Ratio: 0.618 Gaps: 20  
 Percent Similarity: 48.182 Percent Identity: 20.455  
 alignment\_block:  
 US-09-303-518D-131 x A84943 ..  
 Align seg 1/1 to: A84943 from: 1 to: 473  
 127 ATGAAATCAAGCAAGGTGAAGCCGTCAAAAGGCGCAA...GTGCTGTT 173  
 20 leuargValLysIleAsnGluLysValLeuArgGlyGlnProLeuIlePh 36  
 174 TGAAGCAAAAGAAATCCGGCGGTAGTATTACTCGCGCGCTTCAGCA 223  
 36 eSerAspPheAsn.....ValProValHisAlaProThrSerGlyL 51  
 224 AAATCGCCCTATTACCGTGGCGGAAAGCGGTACTTCAGTCAGTCGTG 273  
 51 eulleGluAsnIleCysPheAsnSerAspSerIleLysLysAsnIleLys 67  
 274 ATTCCGCTGAAGGC.....AACGACGA 296  
 68 lleValIleSerProAspTyrLeuAspGlnTrpIleArgLeuAsnProI 84  
 297 ATTCGAGTTCGACGCTAGCTACCTGACGCTGCGCAAAATTCAGCAGC 346  
 84 eLysAspTyrLysLysTyrAlaProGluLysLeuIleLys..... 98  
 347 AAAAAGTGGCGCCCAACCTGATCAATCAATCAGGCTTATGGACTCGCTCGC 396  
 99 .....IleHisGlnSerGlyVal...ValGlyLeuGly 108  
 397 ACCCGTCGCTTC.....ACCAAAATCCCTGCGGTAGATCCGCGAGCC 437  
 109 GlyGlyGlnPheProSerSerLysLysIleIlePheSerIleAsnArgAl 125  
 438 GTTCGCGCATCTCGTCAATCGGTGACACCAATCGCTGGCTGCGGACC 487  
 125 aHisThrLeuIleValAsnAlaValGluSerGluProTyrIleThrSera 142  
 488 CTACGCTCATCATCAAGAACGCGCGCAAGACTTCAACGCGCGCTGTG 537  
 142 spAsnCysLeuIleTyrAsnHisIleSerGluIleLeuIleGlyCysLys 158  
 538 GTATTGCGCGCTGACGACGACGTAATCCATCTGTGTAACGACGACG 587  
 159 ileIleCysTrpIleThrLysIleLysThrValLeuIleAlaIleGlnG 175  
 588 CGCAGAGTCCGCTCT.....GAAATGCTG 613  
 175 uAspAsnIleGlnSerIleSerLysIleGlnHisLeuIleLysAsnLys 192  
 614 CCAATATCGAACAACATGAATTTGGCGCGCGCATCTCGCGCG..... 657  
 192 erLeuPheLysIleCysIleIleLysLysLysTyrProAlaGlySerSer 208  
 658 .....TTGAGTGGCAGCAGCATT.....CA 677  
 209 LysValLeuValLysSerLeuThrGlyLysGluValProHisGlyLysH 225  
 678 TTTCATCGAGCAGCTCGCGCGCAATAAACCGGTGTGGACCATCAATTATC 727  
 225 sSerIleAsp.....TyrLeuIlePheAsnValA 237  
 728 AAGACGTGATTGCTATCGGAGCTTGTTCGTACACGCGCGCTGCAATACC 777  
 237 laThrIlePheSerIleLysArgAlaIleIleAsnGlyLysProLeuThr 253  
 778 GAGCGCTGTTGCTGGCGGCTGCAAGTCAACAAACGCGCGCTT 827  
 254 GluArgValValThrLeu.....MetSerAspLysAsnLeuLe 266

828 CGGTACCGTTTGGGTGCGAAGGTGTCTCAA.....CTTA 862  
 266 uSerGlyAsnPheTrpValArgIleGlyThrProIleLysTyrPheLeuT 283  
 863 CGCGCGCGCAATTTGTTGACGCGGCAACCGCGTATTTCCGTTCCGTTA 912  
 283 hrSerAsnLysLeuLysGlnSer.....PheIleAlaSerValTyr 296  
 913 TTGAACGCTGCTGATTCACAAAGCGCGCATGATTATTGGGAGCGCTACCA 962  
 297 LeuGlyGly.....ProPheMetGlyLysLysI 306  
 963 CAATCAGATTTCGTTATCGAAGAGCGCGCAGCAAGAGAGCTGTTCGGGT 1012  
 306 eAsnAsnLeuAsn..... 310  
 1013 GGGTTCGCGCGCGGACCAATATCTCATCATCAGCGCCACTCTCGGC 1062  
 311 .....HisSerIleLeuLysLysThrAsnSer 319  
 1063 CATTTCTTAAATAAACTCTTCAAGTTCACGACAGCGCGTCAACGCGCG 1112  
 320 IlePheIleThrHisLysLysGluLysAsnGluSerIle.....Se 333  
 1113 CGACCGCGCATGATACCGCATCGGCACCTATGAGCGCGTAAATCGCGTTGG 1162  
 333 rGluLysThrCysIleArgCysGlyTyrCysSerTyrValCysProValA 350  
 1163 ACATCTCTGCTACCTGCTT.....TTGCGCGATTAAATCGTCGCGC 1203  
 350 snLeuLeuProGlnGlnLeuTyrTrpTyrIleLysAsnLysAsnHisVal 366  
 1204 GATACCGACAGCGCGCGAGCTTTGGTGTGCTTGGATTGGACGAGAAAGA 1253  
 367 GlnThrLysLysHisTyrValLeuAspCysIleGlu..... 378  
 1254 CTCTGCTTTGTCAGCTTCGTCTCGCGCGCAATACGAATACGCGCCCGC 1303  
 379 .CysLysAlaCysGluLysValCysPro.....SerTyrIleProL 392  
 1304 TGTTCGCGCAAGTGTCTGAA 1323  
 392 euValLysTyrPheIleGln 398  
 seq\_name: pir2:A40670  
 seq\_documentation\_block:  
 nuclear envelope protein POM 121 - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 05-Nov-1999  
 C:Accession: A40670  
 R:Hallberg, E.; Wozniak, R.W.; Blobel, G.  
 J. Cell Biol. 122, 513-521, 1993  
 A:Title: An integral membrane protein of the pore membrane domain of the nuclear enve  
 A:Reference number: A40670; MUID:93328754  
 A:Accession: A40670  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1199 <HAL>  
 A:Cross-references: GB:221513; NID:g396746; PIDN:CAA79725.1; PID:g396747  
 P:803-807, 845-849, 956-960, 1010-1014, 1047-1051, 1076-1080/Region: pentapeptide motif (X

alignment\_scores:  
 Quality: 129.50 Length: 478  
 Ratio: 0.576 Gaps: 20  
 Percent Similarity: 47.071 Percent Identity: 22.385  
 alignment\_block:  
 US-09-303-518D-131/rev x A40670 ..  
 Align seg 1/1 to: A40670 from: 1 to: 1199



C:keywords: calcium binding; glycoprotein; transmembrane protein  
F1-21/Domain: signal sequence #status predicted <Sig>  
F22-1802/Product: HKRI protein #status predicted <Mat>  
F1483-1508/Domain: transmembrane #status predicted <TM>  
F1645-1656/Domain: calcium binding #status predicted <CAL>

alignment_scores:		
Quality:	128.00	Length: 467
Ratio:	0.554	Gaps: 24
Percent Similarity:	49.465	Percent Identity: 23.340

```
alignment_block:
MS-09-303-518D-131/rev x S69703
```

Align seq 1/1 to: S69703 from: 1 to: 1802

11337	TCCTTCCAAATGGTTTCCAGCACACTTTCGCGAACACAGCGGCCTATTTCGTA	1288
	:::::     :::	
390	SerAspThrValValSerSerSerThrAsnThrValProTyrSerSe	406
1287	TTTCCCGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGCCCAATT	1238
:::	:::    :::	
406	rVal...HisSerThrPheValHis...AlaThrSerSerSerThrTyri	421
1237	CCAAGCAACCACAAAGCTGCGCGTGTCGGGTATCGCGACGATTAAATCG	1188
:::	:::    :::	
421	leSerSerSerLeuTyiSerSerProSerLeuSerAlaSerValSerSer	437
1187	CGCAAAGCAGGTAGGCAGGATGTCMACGGCATTTACGCCCTCATAGT	1138
:::	:::    :::	
438	HisPheGlyValAlaProPheProSerAlaTyriIleSerPheSerSerVa	454
1137	GCCGATC.....GGTACCATGGCGGTGCGCGCGTGTGACGGGTG	1097
	:::    :::	
454	IProValAlaValSerSerThrTyiThrSerSerProSerAlaSerValV	471
1096	TCGTGAACCTTGAAGAGTTGTTTTTAGGAAATGCCGAGAGTGGTCGC	1047
	:::    :::	
471	aIvalProSerAlaTyiAlaSerSerProSerValProValAlaValSer	487
1046	GTGATGGAGTATTGTCGGGTGCGGCCA.....	1017
	:::    :::	
488	...SerThyTyiThrSerSerProSerAlaProAlaIleSerSerTh	503
1016	.....ACCAGCGCAACAGCTCTTCTGCTGCGGCTTCTTCGATAACGG	974
:::	:::    :::	
503	rTyiThrSerSerProSerAlaProValAlaValSerSerThyTyiThrS	520
973	AAATCTGATTGTGTAGCGTCCCAAAATATCATGCGCGCTTGTGCATC	924
:::	:::    :::	
520	eSer.....ProSerAlaProAlaIle	528
923	....GCACGTTCAATACCGAACCGCAA.....ATCAGCGGTTCTTCGCG	883
:::	:::    :::	
529	SerSerThyTyiThrSerSerProSerAlaProValAlaValSerSerTh	545
882	GPfCAACCAATTCGCGGGGGTGAAGTTGAGACACCTTCGCACCCCAAAACGG	833
:	:     :::	
545	rTyiThrSerSerProSer.....AlaProAlaAlaI	556
832	THAGCAAGAGCGCGGTGTTGTACTTGCAGGCGG.....CCCAGGCA	789
:::	:::    :::	
556	le.....SerSerThyTyiThrSerSerProSerAlaProValAla	569
788	ACCAGCGCTCGGTATTCAGACGGCTGTTACGAACAACGTCGATAGC	739
	:::    :::	
570	ValSerSerThyTyiThrSerPro.....SerAlaProValAl	583
738	AATCAGCTCTTGATATATGATGTCACACGCTTTTATTTCGGCGCG.....	693
	:::    :::	
583	alleSer.....ThyTyiThrSerSerProSerV	594

692 .....ACTGGCTCGATGAATGAATGTCG 566  
594 alProValAlaValSerSerThrTyThrSerSer ..... 605  
668 GTGCCACTCAAGCGCGGAGGA.....TGCGGGCCGCCAAA 634  
606 ...ProSerAlaProAlaAlaIleSerSerThrTyThrSerProSe 621  
633 TTCATGTGTTTCGATPATTGGCAGCATTTTCAGACGCAGCTCTGGCCCTG 584  
621 rAlaProValAlaValSerSerThrTyThrSerSerProSerAlaProA 638  
583 CTGCCTTACACATGGATTTCAGTTCG.....GTC 552  
638 laAlaIleSerSerThrTyThrSerSerProSerValProValAlaVal 654  
551 AGCGGCTCAATPACCAACAGCCGCGTTTGAAGTCTTCGGCGGCTTCCTT 502  
655 SerSerThrTyThrSerSerPro.....SerAlaProAlaIle 668  
501 GATGATGACCGTAGGGTCGGCAGCGAGGATGGTGTCTCATCGCATGA 452  
668 eSerSerThrTyThrSerSerProSerValProValAlaValSerSerT 685  
451 CGAAGATGCGCAACGGCTCTACGTCAGCGAGGATTTGCTGTACAGGA 402  
685 hrTyThrSerSerProSerAlaProAlaA.....IleSerSer 698  
401 CGGGTGGCAGCGCAGTCGATAGCCTGATGATCAGGTTCGGCGGCAC 352  
699 ThrTyThrSerSerProSerAlaPro.....ValAlaValSerSerTh 713  
351 TTTTTCGCTGCTCAATTTTGGCCGCGCTTCAGGTACGTAGCTTCGAAC 302  
713 rTyThrSerSerProSerAlaProAlaA.....IleSerSerT 727  
301 CGATTTCGTCGTTCGCTTCACGGCAATCAGCTGACTGACTGAAGTCGCGC 252  
727 hrTyThrSerSerProSerAlaProValAlaValSerSerThrTyThr 743  
251 TTTTCGCCACGGTGATACGCGCGATTTCCTCAAGCCGGCGCAGTAA 202  
744 SerSerProSerAlaProAlaAlaIle.....SerSerThrTyThr 757  
201 TACTACGCCCGGATTCTTTTGTCTTCAACAGCAGCATTCGGCCTTTTGA 152  
757 rSerSerProSerAlaProValAlaValSerSerThrTyThr.....T 771  
151 CGGCTTCACCTTCCTTGATTTCATTCAGGGGCGCATCCCGCATATTCT 102  
771 hrSerSerProSerAlaLeuValValLeuSerSerThrSerSerSer 787  
101 .....TCGGCAAGCAACGCGATTCCTCGGTAAATGCGCGG 70  
788 ProTyAspIleValTyThrSerProSerThrPheAlaAlaIleSerSerGl 804  
69 G 69  
804 I 804

seq name: pir2:S39893

seq documentation\_block:

seq\_documentation\_block: rnfc protein - Rhodobacter capsulatus

rnfc protein - Rhodobacter capsulatus  
C. species: Rhodobacter capsulatus

C;Species: Rhodobacter capsulatus

C;Date: 27-Jan-1995 #sequence\_rev:

C;Accession: S39893

R;Schmehl, M.; Jahn, A.; Meyer zu

Mol. Gen. Genet. 241, 602-615, 1995

**A:Title:** Identification of a new

A;Title: Identification Of a New  
A:Reference number: S39892; MUID:

A;Reference number: S39892; MUID:  
A;Accession: S39893

A;Accession: S39893





A: Experimental source: adult testis; clone DKF2p434C196  
 R: Poustka, A.; Wellenreuther, R.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
 submitted to the Protein Sequence Database, October 1999  
 A: Reference number: Z21540  
 A: Accession: T34549  
 A: Molecule type: mRNA  
 A: Residues: 262-580 <POU1>  
 A: Cross-references: EMBL:AL122069; NID: g6102864; PIDN: CAB59245.2; PID: g7018420  
 A: Experimental source: adult testis; clone DKF2p434B0635  
 R: Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
 submitted to the Protein Sequence Database, September 1999  
 A: Reference number: Z18723  
 A: Accession: T17264  
 A: Molecule type: mRNA  
 A: Residues: 262-580 <POU2>  
 A: Cross-references: EMBL:AL117481; NID: g5911958; PIDN: CAB55954.1; PID: g5911959  
 A: Experimental source: adult testis; clone DKF2p434B061  
 C: Genetix:  
 A: Note: DKF2p434C196.1; DKF2p434B0635.1; DKF2p434B061.1

alignment\_scores:  
 Quality: 121.50 Length: 530  
 Ratio: 0.557 Gaps: 23  
 Percent Similarity: 41.132 Percent Identity: 23.019

alignment\_block:  
 US-09-303-518d-131/rev x T43481 ..

Align seg 1/1 to: T43481 from: 1 to: 580

```

1336 CCTCTCAATGGTTTCAGCACTTCGCGAACAGCGCGGCGATTCGTAT 1287
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
29 ProSerArgAlaSerProThrArgMetProSerArgAlaSerLeuLysMe 45
1286 TTGCCCGGGGACAGAGTGCACAAAG.....CGAGGTCTTCTTC 1246
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
45 tThrProPheArgAlaSerLeuThrLysMetGluSerThrAlaLeuLeuA 62
1245 GTCCAATTCAC..... 1235
|| :|||
62 IgThrLeuProArgAlaSerLeuMetArgThrProThrArgAlaSerLeu 78
1234 .....AGCAA 1230
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
79 MetArgThrProProArgAlaSerProThrArgLysProProArgAlaSe 95
1229 CCGAAAG...CTGCGGCGTGTGGTATCCCGACGATTAATTCGGCAA 1183
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
95 rProArgThrProSerArgAlaSerProThrArgArgLeuProArgAlaS 112
1182 AAGCAAGTAGCGAGGATGTCCAGCGCATTCGCGCTCATAAGTGCAGA 1133
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
112 erPro.....MetGlySerProHis.....ArgAlaSerProMetArg 124
1132 TCGGTACCATGGCCGCGTGCAGCGGTGACGGCTGTCGTAACCTGAAG 1083
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
125 ThrProPro.....ArgAlaSerProThr..... 132
1082 AGTTTGTTTTAGGAATGCCGAGAGTGGTGGCGGTGATGAGTATT 1033
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
133 .....GlyThrProSerThrAlaSerProThrGlyT 143
1032 GTCCGGCTGGCGCAACCGACGCAACAGCTCTTTGCTGGCGCTCTT 983
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
143 hrProSerSerAlaSerProThrGlyThrProPro..... 154
982 CGATAACGAAATCTGATGTGGTAGCGTCCCAATTAATCATGCGCCCT 933
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
155 .....ArgAlaSerProThrGlyThrProProArgAlaThr 169
932 TGTGCAATCGACCGTTCATA.....CCGAACCGGAATCA.... 896
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

169 gSerProSerThrAlaSerLeuThrArgThrProSerArgAlaSerLeuT 186
895 .....CGCGGTGT 887
|||
186 hrArgTPrProProArgAlaSerProThrArgThrProProArgGluSer 202
886 CCGCGTCAACCAATCCCGCGGTAAAGTTGAGACACCTTC.....GCACC 841
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
203 ProArgMetSerHisArgAlaSerProThrArgThrProProArgAlaSe 219
840 CAAAACGGTACGCAAGAGGCGGTTTGTACCTGCGAGCCGCCCAAGG 791
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
219 rProThrArgArgProProArgAlaSerProThrArgThrProProArg 236
790 CAACACCGCTCGCTATTTCAGACGGCTGTTCAGAAACACCTCCGATA 741
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
236 LuSerLeuArgThrSerHisArgAlaSerProThrArgMetProProArg 252
740 GCAATCAGCTTCTGATAATTTGATGCTCCACACGGTTTTTATTCGGCGC 691
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
253 AlaSerProThr.....ArgArgProProArgAlaSerProTh 265
690 TGGCTCGATGAATGATGCTGCCACTCAAGCCGCGAGGATGCGGG. 642
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
265 rGlySerProPro.....ArgAlaSerProMetThrProProArgAlaSerP 281
641 .....CGCCCAAAATTCATGT 627
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
281 roArgThrProProArgAlaSerProThrThrProSerArgAlaSer 297
626 GTTCGATATTGGCAGCATTTTCAGAGCGGACGCTCGGCTGCTGCTTT 577
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298 LeuThrArgThrProSerTrpAlaSerProThrThrProSerArgAl 314
576 ACACACATGGATTTACGTCGTCAGCGGCTCAATACCAACAGCGCCG 527
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
314 aserLeuMetLysMetGluSerThrValSerIleThrArgThrProA 331
526 GTTTGAAG.....TCT 516
|| :||
331 gAlaSerProThrGlyThrProSerArgAlaSerProThrGlyThrPro 347
515 TCGCGCGCTTCTTTCATGATGACCGTAGGTCGCGCAGCCAGGATGGT 466
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
348 SerArgAlaSerLeu.....ThrGlySerProSerArgAlaSerLeuTh 362
465 GTCCATCGATTGACGAAGATGGCGACGCTCGGCATCTACGCGAGGA 416
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
362 rGlyThrProSerArgAlaSerLeuIleGlyThrProSerArgAlaSerL 379
415 TTTTCTGACAGCGGCGGTGGAGCGGAGCGAGTCCATACGCTGATGATC 366
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
379 euIleGlyThrProSerArgAlaSerLeuThrGlyThrProProArgAla 395
365 AGTTTGGCGCGCACTTTTCGCTGCTCAATTT..... 333
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
396 SerLeuThrGlyThrSerSerThrAlaSerLeuThrArgThrProSerAr 412
332 .GCCAGCGCTCAGGTACGTTCGAACTCGAATTCGTCGTCGTCGTCCTT 284
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
412 gAlaSerLeuThrArgThrGlnSerSerSerSerLeuThrArgThrProS 429
283 CAACGGCAATCAGCTGACTGACTGAGTACGCTTTTCGCGCAGGTGAATA 234
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
429 erMetAlaSerLeu.....ThrArgThrProProArg..... 439
233 GCGGGGATTTTG.....CCTGAAGCGCGCGAGTAAATACCTACGCC 193
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
440 AlaSerLeuThrArgThrProProArgAlaSerLeuThrArgThrProPr 456
192 CGGATTCCTTTTCTTCAACAGCAGCTGGCCTTTTTCAGCGCTTCAC 143
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
456 oArgAlaSerLeuThrArgThrProProArgAlaSerLeuThrArgThrP 473

```



```
seq_name: pir2:F85769
382 rLeu 383

seq_documentation_block:
probable membrane protein 22636 [imported] - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Dec-2001
C:Accession: F85769
R:Perina, N.J.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimantanta, E.; Potamouisis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: F85769
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-740 <STO>
A:Cross-references: GB:AE005174; NID:gl2515618; PIDN:AAG56618.1; GSPDB:GN00145; UWGP:226
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
C:Superfamily: unassigned ferredoxin 2[4Fe-4S]-related proteins; ferredoxin 2[4Fe-4S] h
alignment_scores:
Quality: 121.00 Length: 319
Ratio: 0.742 Gaps: 14
Percent Similarity: 51.097 Percent Identity: 22.571

alignment_block:
US-09-303-518D-131 x F85769
Align seg 1/1 to: F85769 from: 1 to: 740
100 GAAGATATGTGGCATGCGCCCTCGATGAAATCAAGGAAGGTGAAGC 149
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
48 LysGlnHisIleGlyAlaGluGlyGluLeuCysValSerValGlyAspLy 64
150 CGTCAAAAGCCCAAGTGTGTTGAAGACAAAGAAATCCGGCGCTAG 199
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
64 sValLeuArgGlyGlnProLeuThrArgGlyArgGlyLysMetLeuPro 81
200 TATTACTGCGCGCTTCAGGAAATCGCGCTATT..... 237
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
81 alHis....AlaProThrSerGlyThrValThrAlaIleAlaProHisSer 96
238 .....CACCGTGGCGAAAGCGCGTACTTCAGTCAGTCGTGATGCGGT 281
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
97 ThrAlaHisProSerAlaLeuAlaGluLeu...SerValIleIleAspAl 112
282 TGAAGGCAACGCAATCGAGTTCGAACGCTAGCTACCT...GAAGCGC 328
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
112 aspGlyGluasp.....CysTrpIleProArgaspGlyT 124
329 TGGCAAAATTCAGCGGAAAGTGGCGCCCAACCTGATTCAA..... 372
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
124 rPalaAspTyrArgSerArgArg...ArgGluGluLeuIleGluArgile 139
373 .....TCAGGCTTATGGACTGGCT 392
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
140 HisGlnPheGlyValAlaGlyLeuGlyAlaGlyAlaGlyPheProThrGlyVa 156
393 TCGCACCGCTCGGTTCAGCAAAATCCCTGCGGTAGATGCGGAGCGGTTCG 442
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
156 l.....LysLeuGlnGlyGlyGlyAspLysIleGluT 167
443 CCATCTTCGTCATGCGATGACCAACCAATCGCGTGGTGGCGACCCCTACG 492
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 hrLeuIleIleAsnAlaAlaGluCysGluProTyrIleThrAlaAspAsp 183
493 GTCATCATCAAGAAGCGCGGAGAGACTTCAACGCGCGCTGTGGTATT 542
```

```
184 ArgLeuMetGlnAspCysAlaAlaGlnValValGluGlyIleArgileLe 200
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
543 GAGCGCGCTGACCGAA..... 558
1:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
200 uAlaHisIleLeuGlnProArgGluIleLeuIleGlyIleGluAspAsnL 217
559 .....CGTAAATCCATGTGTGAAGCAGCGCGCAGACGACGTCGCGTCT 603
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
217 ysProGlnAlaIleSerMetLeuArgAlaValLeuAlaAsp..... 230
604 GAAATGCTGCCATATATGCAACACATGAATTGGCGCGCGCATCTGTC 653
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
231 .....SerHisAspIleSerMetArgValIleProThrLysTyrProSe 245
654 CGGC.....TTGAGTGGCGACGACATTCATT 679
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
245 rGlyGlyAlaLysGlnLeuThrTyrIleLeuThrGlyLysGlnVal.... 260
680 TCATCGACCGCATGCGCGCAATAAAACCGTGTGGACCATCAATATCAA 729
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
261 .....ProHisGlyGlyArgSerSerAspIleGlyValLeuMetGln 274
730 GACGTG.....ATTGCTATCGGACGCTTGTTCGTAAACAGCGCGTCT 770
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
275 AsnValGlyThrAlaTyrAlaValLysArgAlaValIleAspGlyGluPr 291
771 GAATACCGAGCGCGTGTTCGTTGGCGCGCTCAAGTCAACAAACCGC 820
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
291 olleThrGluArgValValThrLeuThrGlyGluAlaIleAlaArgProG 308
821 GCCTCTGCGTACCGTTCGTTGGTGGAGGTGTCTCAACTT.....ACC 864
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
308 lYAsnValTrpAlaArgLeuGlyThrProValArgHisLeuLeuAsnAsp 324
865 GCGCGCGATTTGTTGACGCGGACACCGCTGATTCGCGTTCGGTATT 914
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
325 AlaGlyPheCysProSerAlaAspGlnMetValIleMetGlyGlyProLe 341
915 GAACGCT 921
341 uMetGly 343
seq_name: pir2:B90921
seq_documentation_block:
probable membrane protein [imported] - Escherichia coli (strain O157:H7, substrain RI
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 14-Dec-2001
C:Accession: B90921
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and 9
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: B90921
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-772 <HAY>
A:Cross-references: PIDN:BA000007; PIDN:BA035761.1; PID:gl3361805; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
C:Superfamily: unassigned ferredoxin 2[4Fe-4S]-related proteins; ferredoxin 2[4Fe-4S]
alignment_scores:
Quality: 121.00 Length: 319
Ratio: 0.742 Gaps: 14
Percent Similarity: 51.097 Percent Identity: 22.571

alignment_block:
US-09-303-518D-131 x B90921
```

Align seg 1/1 to: B90921 from: 1 to: 772

```
100 GAAGAAATATGTCGGCGATCGCCCTCGATGAAATCAAGGAGGTGAAC 149
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
48 LysGlnHisIleGlyAlaGluGlyGlnCysValSerValGlyAsp 64
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
150 CGTCAAAAGCCCAAGTCCTGTTTGAAGACAAAGAAATCGGCGGTAG 199
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
64 sValLeuArgGlyGlnProLeuThrArgGlyLysMetLeuProV 81
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
200 TATTTACTCGCGCGCTTCAGCAAAATCGCGGTATT..... 237
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
81 alHis...AlaProThrSerGlyThrValThrAlaIleAlaProHisSer 96
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
238 .....CACCGTGGCGAAAGCGCTACTTCAGTCAGTCGATCGCGT 281
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
97 ThrAlaHisProSerAlaLeuAlaGluLeu...SerValIleIleAsp 112
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
282 TGAGGCAACGACGAAATCGAGTTCGAACGCTAGTACT...GAAGCGC 328
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
112 aAspGlyGluAsp.....CysTrpIleProArgAspGlyT 124
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
329 TGGCAAAATTCAGGACGCAAAAGTGGCCGCCAACCTGATCAA... 372
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
124 rPalaAspTyrArgSerArg...ArgGluGluLeuIleGluArgIle 139
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
373 .....TCAGGCTTATGGACTGCGCT 392
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
140 HisGlnPheGlyValAlaGlyLeuGlyGlyAlaGlyPheProThrGlyVa 156
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
393 TCGCACCGCTCGTTCAGCAAAATCCCTGCGGTAGATGCCGAGCGGTGC 442
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
156 l.....LysLeuGlnGlyGlyAspLysIleGluT 167
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
443 CCATCTTCGTCAATGCGATGACACCAATCCGCTGGCTGCGGACCCCTACG 492
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
167 hrLeuIleIleAsnAlaGluCysGlnProTyrIleThrAlaAsp 183
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
493 GTCATCATCAAGAGCGCGAAGACTTCAAAAGCGCGCTGTGGTATT 542
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
184 ArgLeuMetGlnAspCysAlaAlaGlnValAlaGluGlyIleArgIle 200
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
543 GAGCGCGCTGACCGAA..... 558
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
200 uAlaHisIleLeuGlnProArgGluIleLeuIleGlyIleGluAspAsnL 217
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
559 .....CGTAAATCCATGTCTGTAAGCAGCAGCGCGCAGCTGCCGTCT 603
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
217 ysProGlnAlaIleSerMetLeuArgAlaValLeuAlaAsp..... 230
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
604 GAAATGCTGCCAATATCGAAACACATGAATTTGGCGCGCGCATCTGCG 653
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
231 .....SerHisAspIleSerMetArgValIleProThrLysTyrProSe 245
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
654 CGCG.....TTGAGTGGCAGCCACATTCATT 679
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
245 rGlyGlyAlaLysGlnLeuThrTyrIleLeuThrGlyLysGlnVal.... 260
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
680 TCATCGAGCGAGTCGCGCGGAATAAAACCGTGTGGACCATCAATTATCAA 729
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
261 .....ProHisGlyGlyArgSerSerAspIleGlyValLeuMetGln 274
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
730 GAGCTG.....ATTGCTATCGAGGTTCGTAACAGCGCGTCT 770
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
275 AsnValGlyThrAlaTyrAlaValLysArgAlaValIleAspGlyGluPr 291
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
771 GAATACCGAGCGGTGTTGCTTGGCGCGCTCGCAAGTCAACAACCCG 820
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
291 oileThrGluArgValValThrLeuThrGlyGluAlaIleAlaArgProG 308
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
821 GCCTCTGGGTACCGTTTGGGTGGCAAGGTGTCTCAACTT.....ACC 864
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
308 lyAsnValTrpAlaArgLeuGlyThrProValArgHisLeuLeuAsnAsp 324
865 GCCGGCGAATTTGGTTGACGGGCAACACCGGTGATTTCCGGTTCGGTATT 914
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
325 AlaGlyPheCysProSerAlaAspGlnMetValIleMetGlyGlyProLe 341
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
915 GAACGGT 921
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
341 uMetGly 343
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
seq_name: pir2:I52257
seq documentation_block:
episalin - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
C:Accession: I52257; I65210
R:Vos, H.L.; De Vries, Y.; Hilkens, J.
Biochem. Biophys. Res. Commun. 181, 121-130, 1991
A:Title: The mouse episalin (Muc1) gene and its promoter. Rapid evolution of the rep
A:Reference number: I52257; MUID:92068178
A:Accession: I52257
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-631 <RES>
A:Cross-references: GB:M77226; NID:g199835; PIDN:AAA39754.1; PID:g199837
A:Accession: I65210
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-631 <RE2>
A:Cross-references: GB:M84683; NID:g199842; PIDN:AAA39756.1; PID:g199843
C:Genetics:
A:Gene: Muc1
A:Introns: 20/1; 454/3; 472/2; 517/1; 557/3; 607/3

alignment_scores:
    Quality: 119.00      Length: 456
    Ratio: 0.569        Gaps: 22
Percent Similarity: 45.833 Percent Identity: 24.123

alignment_block:
US-09-303-518D-131/rev x I52257 ..

Align seg 1/1 to: I52257 from: 1 to: 631

1337 TCCTCTCAATGGTTCCAGCACTTTGCGCAACACGCGGCGGTATTCGTA 1288
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
38 SerSerSerLeuAlaSerThrThr..... 45

1287 TTTGCCCGGCAGACGCTCCACAAAGCGAGGTCTTCTTCGTCCTCAATT 1238
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
46 .....ThrProValHisSer.....SerAsnSerAspProA 56
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1237 CCAAGCAACCAAGCTCGCGCTGTCGGTATCGCGACGACGATTAAATCG 1188
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
56 laThrArgProGlyAspSerThrSer...SerProValGlnSerSer 71
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1187 CCACAAACGAGTAGGAGGATGTCACCGGCATTACCGCTCATAGT 1138
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
72 ThrSerSerProAlaThrArgAlaProGluAspSerThrThrAlaVa 88
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1137 GCGGATCGGTACCATGCGCGCTGCGCGCTGACGCGCTCTCTGTAAGT 1088
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
88 lIeuSerGlyThr.....SerSerProAlaThrThrAlaProValAsnS 103
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1087 TCAGAGATTGTTTTTAGGAAATGGCGAGAGTGGTGGCGGTGATGGAG 1038
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
103 erAlaSer.....Ser.....Ser 106
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1037 TATTGTGCGCGTGGCGCAACCCAGCCGACAGCTCTTTGTCGCGGCC 988
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
107 ProValAlaHisGlyAspThrSerSerProAlaThrSerLeuSerLysAs 123
```

```
987 TTCTTCGATAACGGAATCTGATGTTGGTAGCGTCCCAAAATATCATCGG 938
||||| : : : : : |||
123 pSerAsnSerProVal.....ValHisSerGlyThrSerSera 137
||||| : : : : : |||
937 CGCCTGTGCAATCGCACCGTTCAATACCGAACCGGAAATCACGCGGTG 888
||||| : : : : : |||
137 laProAlaThrAlaProValasp..... 145
887 TCAGCGCTCAACCAATTCGCGCGGTAGTTGAGACACCTTCGACCCAA 838
||||| : : : : : |||
146 .....SerThrSerProValHisGlyThrSerSerProAla 160
837 AACGGTACCAAGACGCGGTTTGTGACTTCGAGCGCGCCCAAGCAA 788
||||| : : : : : |||
160 aThr.....SerProGlyAspSerThrSerProAspHisSers 175
787 CCACGCGCTCGGTATTACAGACGCGCTTTACGAACAAACGTCGATAGCA 738
||||| : : : : : |||
175 erThrSer.....ProAlaThr....ArgAlaProGluasp 186
737 ATCAGCTCTTGATAATTGATGTCACACGGTTTATTTCGCGCCGACTGG 688
||||| : : : : : |||
187 SerThrSer.....ThrAlaValLeuSerGlyThrSe 197
687 CTCGATGAATGAATGTCGTGCGCCTCAAGCGCGGAGTGGCGCGC 638
||||| : : : : : |||
197 rSer.....ProAlaThrThrAlaProv 205
637 CAATTCATGTTGTCGATATTCGCA..... 612
205 alaspSerThrSerProValAlaHisaspThrSerSerProAla 221
611 ...GCATTTTCAGAC.....GGCAC 595
222 ThrSerLeuSerGluaspSerAlaSerSerProValAlaHisGlyThr 238
594 GTCTGCGCTGCTGCTTTACACATGGATTTTACGTTGCGTCAGCGCGC 545
||||| : : : : : |||
238 rSerSerProAlaThr.....SerProLeuArga 248
544 TCATACCAACAGCGCGCTTGAAGTCTTCGCGGCTCTTTGATGATG 495
||||| : : : : : |||
248 spSerThrSerProValHisSerAlaSerIleGlnAnlleLys 264
494 ACCGTAGGTCGCGACGACG.....GGATTGGTGCATTCGC 457
||||| : : : : : |||
265 ThrThrSerAspLeuAlaSerThrProAspHisasnGlyThrSerValTh 281
456 ATTGACGAAGATGCGCAACGCGCTCGGCA..... 429
281 rThrThrSerAlaLeuGlySerAlaThrSerProAspHisSerGlyT 298
428 .....TCTACGCGACGGATTTCGCTGAACGGAGGGTG 396
||||| : : : : : |||
298 hrSerThrThrThrAsnSerSerGluSerValLeuAlaThrThrProVal 314
395 CGAAGCGCAGTCCATAAGCCTGATTGAATCAGGTTGCGGCGCACTTTTC 346
||||| : : : : : |||
315 TyrSerSerMet.....ProPheSe 321
345 GCTGCTCAATTTCCGACGCTTCAGTACGTACGCTTCG.....AATCGA 299
||||| : : : : : |||
321 rThrThrLysValThrSerGlySerAlaIleProAspHisasnGlys 338
298 TTTGCTGCTGCTTCACAGCGCATCAGACTGACTGAAGTACGCGCTTT 249
||||| : : : : : |||
338 erSerValLeuProthrSerSerValLeuGlySer..... 349
248 TCGCACGGTGAATAGCGCGGATTTTCCTGAACCGCGCGCAATAATAC 199
||||| : : : : : |||
350 .....AlaThrSerLeuValTyrAsnThrSerAlaIleAlaTh 362
```

```
198 TAGCCCGGATTTCTTTTCTTAAACAGCACTTGCCCTTTTGTGACGG 149
||||| : : : : : |||
362 rThrPro.....ValSerAsnGlyThrGlnPro.....s 372
148 CTTACACCTCTCTGATTTTCATCGAGGGCGCATGCCGACATATTTCTTCG 99
: : : : : : : : : : : |||
372 erValProSerGlnTyrProValSerProThrMetalThrThrSerSer 388
98 CCAAGCAACGCGACTTCG 81
||||| : : : : : |||
389 HisSerThrIleAlaSer 394
seq_name: pir2:T05352
seq documentation_block:
hypothetical protein F8B4.120 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
C:Accession: T05352
R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.;
ewes, H.W.; Mayer, K.F.X.; Schueller, C.
submitted to the Protein Sequence Database, February 1999
A:Reference number: Z15409
A:Accession: T05352
A:Molecule type: DNA
A:Residues: 1-857 <BEV>
A:Cross-references: EMBL:AL034567
A:Experimental source: cultivar Columbia; BAC clone F8B4
C:Genetics:
A:Map position: 4
A:introns: 26/3; 45/1; 74/3; 83/1; 122/2; 165/1; 270/2; 307/1; 731/2; 754/2
A:Note: F8B4.120
C:Superfamily: cyclophilin homology
F:6-162/Domain: cyclophilin homology <CYP>
```

```
alignment_scores:
Quality: 118.00 Length: 438
Ratio: 0.621 Gaps: 21
Percent Similarity: 43.379 Percent Identity: 24.429
alignment_block:
US-09-303-518D-131 x T05352 ..
Align seg 1/1 to: T05352 from: 1 to: 857
111 CGGCATGCGCCCTCGATGAAATCAAGAGGTGAACCCGTCAA..... 155
||||| : : : : : |||
448 ArgGluSerProGlySerGluGlyArgHisValArgArgSerPr 464
156 .....AAAGGCCAAGTCTGTTTGAAGACAAAAGAAAT 189
||||| : : : : : |||
464 oThrLysSerValSerArgSerProValArgValLysLysGluArgAspI 481
190 CCGGCGCTAGTATTACTGCGCGCGCTTCAGGCAAAATCGCGCTATTCA 239
: : : : : : : : : : : |||
481 leSerArgSerPro.....SerLysSerLeuSerArgSer 492
240 CCGTGGCGAAAGCGCGTACTTCAGTACGTCTGAT...TGCCGTTGAG 286
||||| : : : : : |||
493 ProLeuArgSerProLysArgValIleSerArgSerProValArgGlyAr 509
287 GCACGACCAAAATCGAGTTCGAACG...CTAGCTACCTGAACGCGTGGCA 333
||||| : : : : : |||
509 gIleAlaArgSerProSerArgSerProValArgSerAlaSerArgGlys 526
334 AAATTGAGCAG..... 344
526 erLeuGlyArgGlyProLeuArgArgSerSerArgSerProSerArg 542
345 .....CGAAAGTCGCGCGCAACCTGATTCAATCAGCGCTTATGA 385
||||| : : : : : |||
543 SerProValArgSerSer.ArgArgSerLeuSerArgSerProIleGlnL 559
```

```

alignment_scores:
  Quality: 117.00      Length: 313
  Ratio: 0.731        Gaps: 13
  Percent Similarity: 51.118  Percent Identity: 22.684

alignment_block:
US-09-303-518D-131 x G64919  ..

Align seg 1/1 to: G64919 from: 1 to: 740

100 GAAGAAATATCTCGCATCGCCCTCGATCAAAATCAAGGAAGGTGAAGC 149
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
48 LysGlnHisIleGlyAlaGluGlyGluLeuCysValSerValGlyAspLeu 64
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
150 CGTCAAAAAGGCCAAGTGTGTTGAAGACAAAAGAAATCCGGCGGTAG 199
   ||| : : ||||| ||| : : : : : : : : : : : : : : : :
64 svalLeuArgGlyGlnProLeuThrArgGlyArgGlyLysMeLeuProV 81
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
200 TATTACTCGCCGGCTTCAGGCAAAATCGCCCTATT..... 237
   || : : ||||| : : ||||| : : ||||| : : ||||| : :
81 alHis...AlaProThrSerGlyThrValThrAlaIleAlaProHisSer 96
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
238 .....CACGGTGGCGAAAGCGCGTACTTCAGTCTCGTGGATTCGGCT 281
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
97 ThrAlaHisProSerAlaLeuAlaGluLeu...SerValIleIleAspAl 112
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
282 TGAAGCAACGACGAAATCGAGTTCGAACGCTACGTACCTGAA..... 324
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
112 aaSpGlyGluAsp.....CysTrpIleProArgAspGlyT 124
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
325 .....CGCTGGCAAATTCAGCAGCGAAAGTGCGCCGCCAACCTGATT 369
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
124 rpAlaAspTyrArgThrArgSerArgGluLeuIleGluArgIleHis 140
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
370 CAATCAGCTTTATGGACTCGCTTCGCACCCGCTCGCTTCAGC..... 411
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
141 GlnPheGlyVal...AlaGlyLeuGlyAlaGlyPheProThrGlyVa 156
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
412 .AAATCCCTGCGGTAGATGCCGAGCGGTTCGCCATCTTCGTCATCGCA 460
   |||| : : : : : : : : : : : : : : : : : : : : : : : :
156 llysLeuGlnGlyGlyAspLysIleGluThrLeuIleIleAsnAlaA 173
   |||| : : : : : : : : : : : : : : : : : : : : : : : :

```



alignment\_scores:  
 Quality: 116.00 Length: 441  
 Ratio: 0.518 Gaps: 19  
 Percent Similarity: 50.794 Percent Identity: 22.222

alignment\_block:  
 US-09-303-518D-131/rev x 147141 ..

Align seg 1/1 to: 147141 from: 1 to: 528

```

1343 CAGCCCTTCCTCTCTCAATG.....GTTTCCAGCACTTTGGCCACAG 1303
||||| ||| ::||| |||
149 GlnProSerSerSerSerSerProProIleSerSerThrVal..... 162
1302 CGGGCGGTAATTGCTATTTCGCCGGCAGCAAGCTGCACAAAGCGAGGT 1253
||| ||| ||| ||| ||| ||| |||
163 .....SerValGlnProSerSerSer.....SerAlaPro 174
1252 CTTCCTTCGTCCAATTCAGCAACCAAGCCTCGCGCTGTCGGTATGCG 1203
||||| ||| ||| ||| ||| ||| |||
174 hrThrSerAlaThrSerValGlnProSerSer.....SerSerSer 187
1202 CCAGCAATTAATCGCGCAAAAGCAAGTAGGCAAGGATGTCACAGCGCAT 1153
||| ||| ||| ||| ||| ||| |||
188 ProProIleSerSerThrValSerValGlnThrSerSerSerSerSerVa 204
1152 T.....ACGCGCTCATAGTCGCATCGGTACCATCGCGCGCGGT 1115
||| ||| ||| ||| ||| ||| |||
204 iProThrThrSerThrThrSerValGlnPro.....SerSerSerSer 219
1114 CGCGCGCGGTGAGCGGTCTGCTGAACCTTGAGAGATTGTTTTTGAAGAA 1065
||| ||| ||| ||| ||| ||| |||
219 erValProThrThrSerAlaThrSerValArgSerSerSerSerSerSer 235
1064 TGCGCGAGAGTGTGTGCGGTGATGGAGTATTGTCGCGCTCGCGCGCAAC 1015
||| ||| ||| ||| ||| ||| |||
236 ThrPro.....IleProSerThrThrSerVa 244
1014 CCAGCGCAACAGCTCTTTCTCGCGGCT...TCATTGATACGGAATCT 968
||||| ||| ||| ||| ||| ||| |||
244 lGlnProSerSerSerSerSerAlaProThrThrSerAlaThrSerVal. 260
967 GATTGTGGTAGCTGCCAAATAATCATGCGCGCTTGTGCAATCGCACCG 918
::|::|::| ||| ||| ||| ||| |||
261 .....GlnProSerSerSerSerThrThrProIleProSerThr 273
917 TTCAATACCGAACCGGAAATACACGCGGTTGTCCGCTCAACCAATTCGCC 868
::|::|::| ||| ||| ||| ||| |||
274 ThrSerValGlnPro.....SerSerSerSerSerAlaPr 285
867 GCGGGTAAGTTGAGACACCTTCGCCACCCAAACGGTACCAAGAGCGCG 818
||| ||| ||| ||| ||| ||| |||
285 oThrThrSerAlaThrSerValGlnProSerSer..... 296
817 GTTTGTTGACTTCAGCGCCCAAGCAACACACGCGTTCGGTATTTCAGA 768
||||| ||| ||| ||| ||| ||| |||
297 .....SerSerProProIleSerSerThrIleSerValGlnPro 310
767 CGGCTGTTTACGAACAACGTCGATAGCAATCACGCTTTGATAATTGAT 718
::|::|::| ||| ||| ||| ||| |||
311 SerSerSerSerSerProThrThrSerThrSer.....ValGI 325
717 GGTCCACACGGTTTTATTCCGCCGACTGGCTGATGAATCAATGTGCG 668
||| ||| ||| ||| ||| ||| |||
325 nProSerSerSerGlySerAlaProThrThrSer.....AlaI 338
667 TGCCACTCAAGCGGAGTAGGCGCGCCCAATTCATGCTGTTTCGATA 618
||| ||| ||| ||| ||| ||| |||
338 hrSerValGlnProSerSerSerSerSerProProIleSerSerThrIle 354
617 TTGCGACATTTTCAGACGGCAGTCTGCCCTGCTGCTTTACACACATG 568
||| ||| ||| ||| ||| ||| |||
355 SerValGlnProSerSerSerSerProThrThrSerThrSerThr 371

```

seq\_name: pir2:147141

seq\_documentation\_block:  
 gastric mucin (clone PGM-2A) - pig (fragment)  
 C:Species: Sus scrofa domestica (domestic pig)  
 C:Date: 21-Feb-1997 #sequence\_revision 21-Feb-1997 #text\_change 03-Nov-2000  
 C:Accession: I47141; S55315  
 R:Turner, B.S.; Bhaskar, K.R.; Hadzopoulou-Cladaras, M.; Specian, R.D.; LaMont, J.T.  
 A:Title: Pig gastric mucin: isolation and characterization of a cDNA clone with a novel  
 A:Reference number: I47141; MUID:94102478  
 A:Accession: I47141  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-528 <T>R>  
 A:Cross-references: EMBL:U10281; NID:g915205; PID:AA48526.1; PID:g915208  
 R:Turner, B.S.; Bhaskar, K.R.; Hadzopoulou-Cladaras, M.; Specian, R.D.; LaMont, J.T.  
 Biochem. J. 308, 89-96, 1995  
 A:Title: Isolation and characterization of cDNA clones encoding pig gastric mucin.  
 A:Reference number: S55315; MUID:95275264  
 A:Accession: S55315  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-528 <T>U>  
 A:Cross-references: GB:U10281; NID:g915205; PIDN:AA48526.1; PID:g915208  
 C:Superfamily: pig submaxillary mucin





```

613 CysArgCysPheSerArgProSerLysSerGlySerArgProGluLysAr 629
971 TTTCCCTTATCGAAGAGCGCGCAGCAAGAGCTGTTCGCTGGGTGCG 1020
    ::: |||||:::|||||::: ||
629 gSerLysSerGlySerArgProGluLysArgSer..LysSerGlySerAr 645
1021 C...CCGACGGCGGACAAATCTCCATCCATCAGCGCACCACCTCTCGGCATTT 1067
    | :|||:::|||||::: |||||
645 gProGluLysArgSerLysSerGlySerArgProGluLysSer 657
1068 CCTAAACAACTCTTCAAGTTCACGA.....CAGCCGTCAAGCGG 1111
    ||| :|||:::|||||::: ||
658 ....LysArgSerLysSerGlySerArgProGluLysArgSerLysSer 672
1112 GCGACGGCGCCATGTGTACCGATCGGCACCTATGAGCGCGTAATGCCGTG 1161
    :|||::: |||||
673 GlySerArgProGluLysArgSerLys 681
1162 GACATCTCGCTACCTTGTGTCGCGGATTTAATCGTGGCGGATACCGA 1211
    :|||::: |||||
682 .SerGlySerArgProGluLysCys.....GlySerAlaMet.... 693
1212 CAGCGCGCAGGCTTTGGGTGCTTGGAAATGGACGAAGAGACC..... 1255
    :|||::: |||||
694 .....LysArgThrGluGly 698
1256 TGCCTTTGTCAGCTTCTGTCGCGGCAATACGAATACGCGCGCTG 1305
    :|||::: |||||
699 GluLysCysAlaArgLysAsnGlyArgPheAsnSerLys.....ArgCy 713
1306 T...TGGCAAAAGTCTGGAAACCA 1327
    | :|||::: |||||
713 sThrCysThrSerValGlyLysPro 721

seq_name: p1r2:T43311

seq_documentation_block:
fatty-acyl-CoA synthase (EC 2.3.1.86) beta chain - fission yeast (Schizosaccharomyces po
N:Alternate names: fatty-acid synthase beta chain
C:Species: Schizosaccharomyces pombe
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C:Accession: T43311
R:Niwa, H.; Katayama, E.; Yanagida, M.; Morikawa, K.
Protein Expr. Purif. 13, 403-413, 1998
A:Title: Cloning of the fatty acid synthetase beta subunit from fission yeast, coexpress
A:Reference number: 222415; MUID:98359895
A:Accession: T43311
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2073 <NIW>
A:Cross-references: EMBL:AB010274; NID:g4115479; PIDN:BAA36384.1; PID:g4115480
C:Genetics:
A:Gene: fasI
C:Superfamily: yeast fatty-acyl-CoA synthase beta chain
C:Keywords: acetyl-CoA; acyltransferase; carbon-oxygen lyase; coenzyme A; fatty acid bio

alignment_scores:
Quality: 114.50 Length: 461
Ratio: 0.556 Gaps: 25
Percent Similarity: 44.685 Percent Identity: 21.475

alignment_block:
US-09-303-518d-131 x T43311 ..
Align seg 1/1 to: T43311 from: 1 to: 2073

217 TCAGGCAAAATCGCGCTATTCACCTGGCGGCAAAAGCGCTACTTCAGTC 266
    :|||::: |||||
948 ThrGlyLysLeu.....MetArgIleGluGluArgPheThrGlnAs 962
267 AGTC.....GTGATGCCGCTTGAAGC.....AACGACGAA 298
    :|||::: |||||

```

```

962 pValGlyLysThrThrLeuIleGluAsnPheGluAspLeuAsnAspProT 979
299 TCGAGTTCCGAACGGCTACCTGAAGCGCTGGCAAAATTTGAGCAGCGAA 348
    :|||::: |||||
979 yrProValAlaAlaArgPheLeuAspAlaTyrProGluAlaSerThrGln 995
349 AAAGTCCCGCCCAACCTGATTCAATCAGGCTTATGAGCTGCGTTCGCAC 398
    :|||::: |||||
996 AspLeu.....AsnThrGlnAspAlaGlnPhePheTyrSerLeuCysSe 1010
399 CCGTCCGCTTCAGCAA.....ATCCCTGCCGTAGAT..... 429
    :|||::: |||||
1010 rAsnProPheGlnLysProValProPheIleProAlaIleAspAspThrP 1027
430 .....GCCGAGCCGCTTCGCC 444
1027 heGluPheTyrPheLysLysAspSerLeuTrpGlnSerGluAspLeuAla 1043
445 ATCTTCGTCAAT.....GCGATGGACACCAATCCGCT 476
    :|||::: |||||
1044 AlaValValGlyGluAspValGlyArgValAlaIleLeuGlnGlyProme 1060
477 GCGTCCCGACCCCTACGGTTCATCAAGAGCGCGCAAGACTTCAAAAC 526
    :|||::: |||||
1060 tAlaAlaLysHisSerThrLysValAsnGluProAlaLysGlu..... 1074
527 GCGCCCTGTTGGTATTGAGCGGCTGACCGACGTAATAATC...CATGTG 573
    :|||::: |||||
1075 .....LeuLeuAspGlyIleAsnGluThrHisIleGlnHisPhe 1087
574 TGTAAAGCA.....GCAGCGCGCAGAC.....GTCCGCTCTGAAATGC 611
    :|||::: |||||
1088 IleLysLysPheTyrAlaGlyAspGluLysLysIleProle..... 1101
612 TGCATATTCGAAACACATGAATTGGCGCGCGCATCTCTGCGGCTGA 661
    :|||::: |||||
1102 .....ValGluTyrPheGlyGlyValProProValAsnValS 1114
662 GT.....GGCAGCGAC 672
    :|||::: |||||
1114 erHisLysSerLeuGluSerValSerValThrGluAlaGlySerLys 1130
673 ATTCAATTCATCGAGCAGCTGCGCGCAATAAACCCTG..... 711
    :|||::: |||||
1131 ValTyrLysLeuProGluIleGlySerAsnSerAlaLeuProSerLysLy 1147
711 ..... 711
1147 sLeuTrpPheGluLeuAlaGlyProGluTyrThrTrpPheArgAlaI 1164
712 .....TGGACCATCAATTATCAAGAC 732
    :|||::: |||||
1164 lePheThrThrGlnArgValAlaLysGlyTrpLysLeuGluHisAsnPro 1180
733 GTGATTGCTATC.....GGACGTTTGTCTGCTAACAGG 764
    :|||::: |||||
1181 ValArgArgIlePheAlaProArgTyrGlyGlnArgAlaValValLysGl 1197
765 CCGTCTCTGAATACCGACCGCTGTCCTTGGCGCGCTGCAAGTCAACA 814
    :|||::: |||||
1197 yLysAspAsn...AspThrValValGluLeuTyrGluThrGlnSerGly. 1212
815 AACCCGCGCTCTTGGGTACCGTTTGGGTGCGAAGGTGTCTCAACTACC 864
    :|||::: |||||
1213 .....AsnTyrValLeuAlaAlaArgLeuSer..... 1221
865 GCCGGGGAATTGGTTGACCGGACAAACCGCTGATTTCCCGTTCGGTATT 914
    :|||::: |||||
1222 .....TyrAspGlyGluThrIleValValSerMetPheGluAs 1234
915 GAACGGTTCGATTGCACAAAGCGCGCAT...CATATTGCGACCTAC. 960
    :|||::: |||||
1234 nArgAsnAlaLeuLysLysGluValHisLeuAspPheLeuPheLysTyrG 1251

```

299 TCAGTTCGAACGCTAGCTACCTGAAAGCGCTGGCAAAATGTAGCAGCGAA 348  
 979 yrProValAlaAlaArgPheLeuAspAlaTyrProGluAlaSerThrGln 995  
 349 AAAGTCGCGCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCAC 398  
 996 AspLeu.....AsnThrGlnAspAlaGlnPhePheTyrSerLeuGly 1010  
 399 CCGTCCGCTTACGAA.....ATCCCTGCGGTAGAT..... 429  
 1010 rAsnProPheGlnLysProValProPheIleProAlaIleAspThrP 1027  
 430 .....GCCGAGCCGCTTCGCC 444  
 1027 heGluPheTyrPheLysLysAspSerLeuTrpGlnSerGluAspLeuAla 1043  
 445 ATCTTCGTCAAT.....CGCATGGACACCAATCGCT 476  
 1044 AlaValValGlyGluAspValGlyArgValAlaIleLeuGlnGlyProMe 1060  
 477 GGCTGCGGACCTACGCTCATCAATCAAGAGCGCGCAAGACTTCAAC 526  
 1060 tAlaAlaLysHisSerThrLysValAsnGluProAlaLysGlu..... 1074  
 527 CGGCGCTTGTGTTATGAGCGCTGACCGCAACGTAATC....CATGTG 573  
 1075 .....LeuLeuAspGlyIleAsnGluThrHisIleGlnHisPhe 1087  
 574 TGTAAGCA.....GCAGGCGCAGAC.....GTGCGCTCTGAAATGC 611  
 1088 IleLysLysPheTyrAlaGlyAspGluLysLysIleProIle..... 1101  
 612 TGCCAATATCAAAACACATGAAATTTGGCGCGCCCATCTGCGCGCTGA 661  
 1102 .....ValGluTyrPheGlyGlyValProValAsnVal 1114  
 662 GT.....GCACGCGAC 672  
 1114 eHisLysSerLeuSerValSerValThrGluAlaGlySerLys 1130  
 673 ATTCAATTCATCAGCAGCGCGCGCAATATAACCGTG..... 711  
 1131 ValTyrLysLeuProGluIleGlySerAsnSerAlaLeuProSerLysLy 1147  
 711 ..... 711  
 1147 sLeuTrpPheGluLeuLeuAlaGlyProGluTyrThrTrpPheArgAlaI 1164  
 712 .....TGGACCATCAATATCAAGAC 732  
 1164 lePheThrThrGlnArgValAlaLysGlyTrpLysLeuGluHisAsnPro 1180  
 733 GTGATTGCTATC.....GGACGCTTGTTCGTACACAGG 764  
 1181 ValArgArgIlePheAlaProArgTyrGlyGlnArgAlaValValLysG 1197  
 765 CCGTCTGAATACCGCGCTGTTGCTTGGCGCGCTGCAAGTCAACA 814  
 1197 yLysAspAsn...AspThrValValGluLeuLysGlnSerGly. 1212  
 815 AACCGCGCTCTTGGCTACCGCTTTGGTGGCGAAGGTGCTCAACTTACC 864  
 1213 .....AsnTyrValLeuAlaAlaArgLeuSer..... 1221  
 865 GCGCGCAATTTGTTACCGCGCAACACCGCTGATTTCGGTTCGTATT 914  
 1222 .....TyrAspGlyGluThrIleValValSerMetPheGluAs 1234  
 915 GAACGTCGCGATTGCACAGGCGCGCAT...GATTATTGGGACGCGTAC. 960  
 1234 nArgAsnAlaLeuLysGluValHisLeuAspPheLeuPheLysTyrG 1251  
 961 .....CACAAATCAGATTTCCTGTTATCGAAGAGCGCGCAGC 996

961 .....CACAAATCAGATTTCCTGTTATCGAAGAGCGCGCAGC 996  
 1251 luProSerAlaGlyTyrSerProValSerGluIleLeuAspGlyArgAsn 1267  
 997 AAAGAGCTG.....TTCGGCTGGGTT.....GCCCGCA 1025  
 1268 AspArgIleLysHisPheTyrTrpAlaLeuTrpPheGlyGluGluProTy 1284  
 1026 GCGGACAAATCTCATCACCGCACACTCTCGCCATTTCCCTAATAA. 1074  
 1284 rProGluAsnAlaSerIleThrAspThrPheThrGlyProGluValThrV 1301  
 1075 .....AACAACTCTTCAAGTTACGACAGCGCTCAACGCGCGCGAC 1116  
 1301 alThrGlyAsnMetIleGluAspPheCysArgThrValGlyAsnHisAsn 1317  
 1117 CGCGCATGTCACCGATCGCACTTATGAGCGCTTAATCGCTTGGAC... 1164  
 1318 GluAlaTyrThrLysArgAlaIleArgLysArgMetAlaProMetAspPh 1334  
 1165 .....ATCCTGCCTACCT 1177  
 1334 eAlaIleValValGlyTrpGlnAlaIleThrLysAlaIlePheProLysA 1351  
 1178 TGCTTTGGCGGATTTA.....ATCGTCGGCGATACCGACGCGCGCAG 1221  
 1351 laIleAspGlyAspLeuArgLeuValHisLeuSerAsnSerPheArg 1367  
 1222 GCCTTGGGTTGCTTGAATTTGGACGAGGAGAGAC 1254  
 1368 MetValGlySerHisSerLeuMetGluGlyAsp 1378  
 seq\_name: pir2:T39207

seq documentation\_block:  
 fatty acid synthase, subunit beta - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000  
 C:Accession: T39207  
 R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Ransperger, U.; Pohl, T.  
 submitted to the EMBL Data Library, September 1999  
 A:Reference number: Z21835  
 A:Accession: T39207  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-2073 <WOO>  
 A:Cross-references: EMBL:AL110469; PIDN:CAB54157.1; GSPDB:GN00066; SPDB:SPAC926.09c  
 A:Experimental source: strain 972h; cosmid c926  
 C:Genetics:  
 A:Gene: SPDB:SPAC926.09c  
 A:Map position: 1  
 C:Superfamily: yeast fatty-acyl-CoA synthase beta chain

alignment\_scores:  
 Quality: 114.50 Length: 461  
 Ratio: 0.556 Gaps: 25  
 Percent Similarity: 44.685 Percent Identity: 21.475

alignment\_block:  
 US-09-303-518d-131 x T39207 ..  
 Align seg 1/1 to: T39207 from: 1 to: 2073  
 217 TCAGCAAAATCGCGCTATTCACCGTGGCGAAGAGCGCTTCTAGTC 266  
 948 ThrGlyLysLeu.....MetArgIleGluGluArgPheThrGlnAs 962  
 267 AGTC.....GTGATTGCGCTTGAAGG.....AACGACGAAA 298  
 962 pValGlyLysThrThrLeuIleGluAsnPheGluAspLeuAsnAspProt 979



1331 TCATGTTTCCAGCACTTTCGCAACAGCGGCGG.....TATTC 1291  
1332 TCATGTTTCCAGCACTTTCGCAACAGCGGCGG.....TATTC 1291  
20 SerThrTyrProAsnThrTyrSerAsnSerAsnProAsnAlaAspAlaSe 36  
1290 GTATTTCGCGGG...CAGACGAGCTGCACAAAGCGAGGTCTTCTTCGT 1244  
36 rSerMetProSerThrAsnThrThrValProGlnThrSerSerSerS 53  
1243 CCATATCCAAAGCAACCAAGCCTCGCGCTGTCGGTATCGCGCAGGATT 1194  
53 erThrSerSerThrThrAlaThrCluSerSerSerGlyThrAla 69  
1193 AATTCGCGCAAAAGC...AAGTAGGAGGATGTCACACGGCATTCGCG 1147  
70 GluSerSerSerThrLysSerAlaThrMetSerGlySerThrThrHl 86  
1146 CTCATAAGTCCGATCGGTACCATGCGCGGTGCGCGCGTTCGACGCTG 1097  
86 sThr.....ThrSerAlaThrAlaSerSerThrAla 98  
1096 TCGTGAACCTTGAGAGATTTCGTTTATAGGAATGGCGGAGAGTGGTGGC 1047  
98 erThrSerThrSerSer..... 103  
1046 GTGATGAGATTCGCGGTGCGCGGCAACCGACCGGACAGCTCTTT 997  
104 .....TyrSerThrSerThrSerThrLysThrThrHl 117  
996 GCTGGCGCTTCGATACGGAATCTGATTGCTGAGTGGTGGTGGC 947  
117 rMetThrGlySerThrSer..... 124  
946 AATCATGCGCGCTTCGCAATTCACCGTTCATACCGAACCGGAATC 897  
125 .....ThrAlaSerAlaAlaProThrSerThr..... 134  
896 AGCGGTTCGCGGTCAACCAATTCGCGGCGGTGAGTTCGAGACACT 847  
135 .....AlaSerThrSerThrSerThrSerThrSerThrSe 149  
846 CGCACCCAAACGGTACGC.....AAGAGCGCGGTTCGTTGACTT 806  
149 rSerThrLysThrThrValThrGlySerThrIleGlyThrAla 166  
805 CGAGCGCGCGCAACGACCGCTCGGTATTCAGA..... 768  
166 erAlaAlaProThrSerThrSerThrSerThrAlaAsnSerAlaSer 182  
767 .....CGGCTTCGCAACCAACGTCGATAGCAATCAGTCTCTG 727  
183 SerThrThrAsnProSerSerGlySerLysProThrAlaMetThrGly.. 198  
726 ATAATGATGTCACACGGTTCGCGCGGTTCGCGCGTTCGATGAAAT 677  
199 .....ThrThrAlaAsnThrSerProSerAlaProThrSer..... 211  
676 GAATGTCGTCCTCAACGCGGAGGATGCGCGCGCAATTCATGT 627  
212 .....ProSerThrThrAsn 216  
626 GTTTCGATATTCGACATTTTCAGACGCGCAGCTCTGCGCTCTGCTTT 577  
217 SerSerThrAlaAlaThrThrSerSerGlySerLysProThr..... 231  
576 ACACATGGAATTTACGTTTCGTCGCGGTCAATACCAACAGCGCGC 527  
232 .....ThrValThrArgThrThrAlaAsn..... 239  
526 GTTTCGATGTCGCGGTTCGTCGCGGTTCGATGATGACCGTACGCGC 477  
240 .....ThrSerSerAlaSerThr.....SerSerAla 249  
476 AGCGGATTCGTCCTCAATTCGATGAGAGATGGCGGAGCGTGGCGCATC 427

938 CCATGATATTATTTGGAGCGCTACCAATCAGATTTCGTTATCAAGAA 987  
542 LuHisMetIleLeuSerIleGlyGlnHisProValLysPheIleLysGlu 558  
988 GCGCGCAGCAAGAGCTGTCGCTGGTGGCGCGCAGCGGACAAATA 1037  
559 GlyLysGluGluLeuValGluLysLeuAlaGluGlyIleGluLysVa 575  
1038 CTCATCAGCGCACCACTCTCGGCATTT..... 1067  
575 lAlaAla..AlaPheTyrProArgProValTrpTyrArgThrLeuAspAl 591  
1068 .CTAAACAACTCTCAAGTT.....CAGCAGCCGCTCAAC 1107  
591 aProThrAsnGluPheArgGluMetProGlyGlyGluAspGluProGlu 608  
1108 GCGCGCAGCGCGCAGCTGATCGGCTGATGAGCGGCTAATGCC 1157  
608 LuArgAsnProMetLeuGlyTrpArgGly.....IleArgArgGly 621  
1158 GTTGGACATCCGCTACCTGCTTTCGCGATTTAATCGTCGCGATA 1207  
622 LeuAspGlnProGluLeuLeuArg...AlaGluPheLysAlaIleLysLy 637  
1208 CCGCAGCGCGCAGGCTTGGGTGCTGGAATGAGACGGAAGACCTC 1257  
637 sValValGluLysGlyTyrAsnAsnIleGly..... 647  
1258 GCTTGTGCGAGCTTCGCTG.....CCGCGCAATAGATACGG 1298  
648 .....ValMetLeuProLeuValSerHisProGluGlnIleArgLysAla 662  
1299 CCGCGTGTTCGCAAGTCTGGA 1322  
663 LysGluIleAlaArgSerValGly 670  
seq\_name: pir2:E86185  
seq\_documentation\_block:  
hypothetical protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: E86185  
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: E86185  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-402 <STO>  
A:Cross-references: GB:AE005172; NID:g2388584; PIDN:AAB71465.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1  
alignment\_scores:  
Quality: 112.50 Length: 438  
Ratio: 0.571 Gaps: 18  
Percent Similarity: 44.977 Percent Identity: 22.831  
alignment\_block:  
US-09-303-518D-131/rev x E86185 ..  
Align seg 1/1 to: E86185 from: 1 to: 402



```

250 Ser.....:::|||||:::|ProThrAsnSerSerThrSe 257
426 TACGGCAGGGATTGTTGCTGAACGCGGGTGCACGCGAGTCCATAAGC 377
257 rThrPro.....|ThrAsnSerSerAlaGlySerLysP 268
376 CTGATTGAATACAGTTGCGCGGCACATTTTCGTGCTCAATTGCGCAGC 327
268 ro.....|ThrThr 270
326 GCTTCAGGTACGTAGCTTCGAACTCGATTTCGTGCTTCCTCAACGGC 277
271 MetThrGlyThrThrThrAsnThrSerSerThrThrThrSerSerAl 287
276 AATCAGCACTGACTGAAGTACGCGCTTTTCGCCACGCGTCAATAGCGGCA 227
287 aserThrThrLysSerSer..... 294
226 TTTTGGCTGAACGCGCGCAGTAATACTACGCGCGGATCTTTTGTCT 177
295 .....SerSerAlaThrAsnSerSerSerGlySerLysProSer 307
176 TCAACAGACACTGGCCCTTTTGGCGCTTCACCTTCCTGATTTCAT 127
308 ThrLeuSerThrThrThrAlaThrThrAlaThrThrSerSerProThrAl 324
126 CGAGGGGCGCATCGCGACATATCTTCGCAAGCAACGCGACTTCGGTAA 77
324 agluProSerThrThrThrAlaSerLysProAlaThrSerSerThrProp 341
76 TGGCGCGCGCGTCA 63
341 roAlaProProThr 345

```

seq\_name: piri:A36904

seq\_documentation\_block:

prochlorophyllide reductase (EC 1.3.1.33) [similarity] - Synechococcus sp. (PCC 7942)

A;Variety: PCC 7942

C;Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C;Accession: S25621; A36904

R;Lileman-Hurwitz, J.; Ronen-Tarazi, M.; Gabai, C.; Hassidim, M.; Schwarz, R.; Kaplan, A. submitted to the EMBL Data Library, August 1992

A;Reference number: S25616

A;Accession: S25621

A;Molecule type: DNA

A;Residues: 1-466 &lt;LIE&gt;

A;Cross-references: EMBL:X67694; NID:g312508; PIDN:CAA47924.1; PID:g46847

R;Price, G.D.; Howitt, S.M.; Harrison, K.; Badger, M.R.

J. Bacteriol. 175, 2871-2879, 1993

A;Title: Analysis of a genomic DNA region from the cyanobacterium Synechococcus sp. strain A36904

A;Reference number: A36904; MUID:93259930

A;Accession: A36904

A;Molecule type: DNA

A;Residues: 388-466 &lt;PRI&gt;

A;Experimental source: PCC 7942

A;Note: sequence extracted from NCBI backbone (NCBIN:131726, NCBI:P:131727)

C;Superfamily: prochlorophyllide reductase chain chlN

C;Keywords: chlorophyll biosynthesis; oxidoreductase; photosynthesis

alignment\_scores:

Quality:	112.50	Length:	471
Ratio:	0.577	Gaps:	25
Percent Similarity:	41.401	Percent Identity:	20.807

alignment\_block:

US-09-303-518D-131 x A36904

Align seg 1/1 to: A36904 from: 1 to: 466

```

7 AAATCAAAAAGGTCTAAATCTGCCCATCGGGGCGAGACGCGAGCAAGT 56
:::|||||:::|
92 GlnIleLysArgAspArgAsn.....ProSerValil 102
57 CATTTATGACGCGCGCGCATTTACCGAGTCCGCTTGTTCGGCGAAGAT 106
:::|||||:::|
102 evalrpileglyThrCysThrThrGluIleIleLysMetAspLeuGlu. 118
:::|||||:::|
107 ATGTGCGCATCGCCCTCGATGAAATCAAGGAAGGTGAAGCGGTCAAA 156
:::|||||:::|
119 .....GlyLeuAlaProLysLeuGluAlaGluIleGlyPro..... 131
157 AAGGCCAAGTCGTTTGAAGACAAAGAATCCGGGCGGTAGTATTATAC 206
:::|||||:::|
132 .....ileValVa 134
207 TGC CGCGCTTACGCAAAATCGCGCTATTTCACCGTGGCGAAGCGCG 256
:::|||||:::|
134 lalaArgAlaAsnGlyLeuAspTyrAlaPheThrGlnGlyGluAspThrV 151
257 TACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATCGAGTTC 306
:::|||||:::|
151 alLeuAlaAlaMetAlaAlaArg..... 158
307 GAACGCTACGTACCTGAGCGCTGGCAAAATTCAGCAGCGCAAAAGTGC 356
:::|||||:::|
159 .....CysProGluAlaAlaThrSerGluAlaAspGlnGlnGluAr 172
357 CCGCAACCTGATTCAATCAGCGCTTATGGACTGCGCTTCGCACCCGTCGT 406
:::|||||:::|
172 gThrAsnAlaIleGln.....ArgLeuLeuGlnP 182
407 TCAGCAAAATCCCTGCGGTAGATGCCGAG..... 435
182 heGlyLysSerProAlaAlaGlnGlnProAlaSerSerLysHisPro 198
436 CCGTTGCCCATCTTCGTCATGCGATGGACACCAATCCGCTGGCTCCGA 485
:::|||||:::|
199 ProLeuIleLeuPheGlySerValProAsp.....ProValAlaThrGl 213
486 CCTTACGTCATCATCAAGAAGCCCGCGAAGACTTCAAAACGCGCTGT 535
:::|||||:::|
213 nLeuThrIle.....GluLeuAlaLysGlnGlyIleT 224
536 TGTATTGACCGCCCTGACCGACGTAATAATCCATGTGTGTAAGCAGCA 585
:::|||||:::|
224 hrValSerGlyTrpLeuProAlaLysArgTyr..... 234
586 GCGCGCAGCGTCCG.....TCTGAAAATGCTGCCAATATCGAAACACA 629
:::|||||:::|
235 ...ThrGluLeuProValIleAlaGluGlySerTyrAlaIleGlyLeuAs 250
630 TGAATTT..... 636
250 nProPheLeuSerArgThrAlaThrThrLeuMetArgArgLysCysL 267
637 .....GCGCGCGCGCATCTCGCGCTTGTGAGTGGCAGCAC..... 672
267 ysValIleGlyAlaProPheProIleGlyProAspGlySerArgAlatr 283
673 .....ATTCAATTCATCGACCGCGTGGC..... 696
284 ileGluLysileCysSerValLeuGluIleGluProGlnGlyLeuAlaGl 300
697 .GCGAATAAAACCGTGTGGACCATC.....AATTATCAAGACGTGATTG 739
:::|||||:::|
300 uargGluAlaGlnValTrpAspSerIleGluAspTyrArgGlnLeuValG 317
740 CTATCGGACGTTTGTTCGTAACAGGC.....CGTCTGAATACC 777
317 luGlyLysGlnValPhePheMetGlyAspAsnLeuTrpGluIleSerLeu 333
778 GAGCGCGTGTTCCTTGGCGGCGCTG.....CAAGTCAACAAACC 818

```







541 ProAsnSerAla.....AspSerLysMetSerPheAsp..... 551  
 237 AATAGCGGCGATTTCCTGTAAGCGCGCGAGTAATACTAGCGCGCGAT 188  
 552 .....LeuProGluLysGlnAspGlyAlaThrSerProGlyA 564  
 187 TCTTTTGTCTTCAACAGCACTTGGCTTTTTCGACGGCTTCACTTCC 138  
 564 laLeuLeuProAlaSerThrThrSerPhePheThrSerAsnProHis 580  
 137 TTGATTTCATCGAGGGCGCATGCGACATATCTTCGCCAAGCAAGC 88  
 581 AspSerLeuValMetAsnThrLeuThrSerPheSerProLeuAsnG1 597  
 87 GACTTCGGTAATGGCGCGCTCATAAATGACTTGTCTCGGTCTGCCC 39  
 597 ncluthrAlaValGluAlaProSerArgArgThrAspAspGluLeuPro 613  
 seq\_name: pir2:C75580

seq\_documentation\_block:  
 adenine deaminase-related protein - Deinococcus radiodurans (strain R1)  
 C:Species: Deinococcus radiodurans  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
 C:Accession: C75580  
 R:White, O.; Eisen, J.A.; Heldelberg, J.F.; Hickey, E.K.; Peterson, R.J.;  
 S.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
 Science 286, 1571-1577, 1999  
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; MUID:20036896  
 A:Accession: C75580  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-376 <WHI>  
 A:Cross-references: GB:AE001863; GB:AE001825; NID:g6460670; PIDN:AAF12376.1; PID:g646067  
 C:Genetics:  
 A:Gene: DRA0268  
 A:Map position: 2

alignment\_scores:  
 Quality: 111-50 Length: 290  
 Ratio: 0.961 Gaps: 15  
 Percent Similarity: 40.000 Percent Identity: 26.897  
 alignment\_block:  
 US-09-303-518D-131 x C75580 ..  
 Align seg 1/1 to: C75580 from: 1 to: 376  
 639 CGGCCCGCA.....TCCTGCGGCTTGAGTGGCAGCGACA 673  
 35 ArgProAlaLeuArgGlyAspSerAlaArgLeuAlaHisAlaHi 51  
 674 TTCATTTCAT.....CGACCCAGTCCGCGAATAAACCGTGTGACC 717  
 51 sHisThrHisPheGlyAlaAlaGlyArgThrGly..... 62  
 718 ATCAATTATCAAGACGTGATTGTCTATCGGAGCTTTGTCTGTAACAGCGCG 767  
 63 .....ThrGlyArgPro 66  
 768 TCTGATATACGAGCGGTGTTGCTTGGCGGCTGCAAGTCAACAAC 817  
 67 AlaAlaGlyProAlaGlyGlnSerGlyAlaGlyCluProArgGva 83  
 818 CGCG...CCCTTCGCTAC.....CGTT 837  
 83 lAlaHisProAlaLeuThrSerGlyGlyArgAlaAspArgAlaGlyArg 100  
 838 TTGGGTGCGAAGGTGCTCAACTACCGCGCGCGAATTGGTTGACGCGGA 887

100 laGlySerAlaAlaGlyAlaArgTyrArgGluLeuGly...ProAla 115  
 888 CAACCGCGCTGATTTCGCGTTCGGTATTGAACGGTGGGATTGCACAGGCG 937  
 116 ArgProArgGlyAspArgProTyrPvalSerArg.....ArgLeuArgAl 130  
 938 CGCATGATTATTGGGACGCTA..... 959  
 130 aAlaAlaArgLeuAlaThrLeuArgGlyAlaGlyAsnLeuArgGlyArg 147  
 960 .....CCACAATCAGATTTCCGTTAT 980  
 147 rgGlySerAlaAlaArgArgArgAspAlaProAla.....ArgTyr 161  
 981 CGAAGAGCGCGCAGCAAGAGCTGTTCGGCTGGGT..... 1016  
 162 ArgArgArgProArg.....AlaArgLeuGlyArgGlyHisLeuAr 175  
 1016 ..... 1016  
 175 gSerAlaArgAlaLeuAlaHisAlaProAspValSerArgProAspArgH 192  
 1017 .....TCGCGCGCAGCGGCAAAATACTCATCACCGCAGCAGCTCTCGS 1061  
 192 isArgAlaCysGlyAlaGlyGlnArgArgProAlaGlyArgArg 208  
 1062 CCATTTCTTAAACAACTCTTCAAGTTCAGCAGCGCTCAACGGCG 1111  
 209 ProLeuArgAlaArgValValGluLeuLeuAspValGlyGlnArgPr 225  
 1112 GCGACCGCGCATGTTACCGATCGGCACATTATCAGCGCGTAATCGGTTG 1161  
 225 oAlaArgArgHisProGlyHisGlnHisSer...AlaArgArgAlaSerG 241  
 1162 GACATCTCGCTACCTT.....GCTTTT 1184  
 241 lyGlyProAlaArgArgGlnArgArgGlyProAlaArgGlyGlySerGly 257  
 1185 GCGCGATTAACTCGTCGGGATACCGACAGCGGCGAGCTTTGGTGTGT 1234  
 258 AlaArgThrAlaArgArgHisArgPro..... 267  
 1235 TGGATTGGACGAAGAAGACCTCGCTTGTGTCAGCTTCGTCGCCGCGGC 1284  
 268 .GlyArgGlyArgArgGlyProArgAlaValAlaProAlaLeuArgGlys 284  
 1285 AATACGAATACGCGCGCT 1304  
 284 er...AspAspArgProAla 289  
 seq\_name: pir2:JC4176

seq\_documentation\_block:  
 Pyruvate,water dikinase (EC 2.7.9.2) - Pyrococcus furiosus  
 A:Alternate names: phosphoenolpyruvate synthetase  
 C:Species: Pyrococcus furiosus  
 C:Date: 10-Sep-1995 #sequence\_revision 27-Oct-1995 #text\_change 21-Jul-2000  
 C:Accession: JC4176  
 R:Jones, C.E.; Fleming, T.M.; Piper, P.W.; Littlechild, J.A.; Cowan, D.A.  
 Gene 160, 101-103, 1995  
 A:Title: Cloning and sequencing of a gene from the archaean Pyrococcus furiosus with  
 A:Reference number: JC4176; MUID:95354939  
 A:Molecule type: DNA  
 A:Residues: 1-817 <JON>  
 A:Cross-references: EMBL:X80819; NID:g967059; PIDN:CAA56785.1; PID:g967060  
 A:Experimental source: Vcl DSM 3638  
 C:Genetics:  
 A:Gene: ppsa  
 C:Function:  
 A:Description: catalyzes the ATP-dependent conversion of pyruvate into phosphoenolpyr  
 A:Pathway: gluconeogenesis

C;Superfamily: Escherichia coli pyruvate,water dikinase; phosphotransferase system enzym  
C;Keywords: phosphoprotein; transferase  
F;66-816/Domain: phosphotransferase system enzyme I homology <Ptl>

alignment\_scores:  
Quality: 111.00 Length: 489  
Ratio: 0.459 Gaps: 24  
Percent Similarity: 49.489 Percent Identity: 21.268

alignment\_block:  
US-09-303-518d-131 x JC4176 ..

Align seg 1/1 to: JC4176 from: 1 to: 817

```

31  CCATCGCGGAGAGCGGAGCAAGTCATTTATGAC..... 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
235  ProvalThrAsnAsnArgAsnGluMetIleAsnAlaSerTirpGlyLe 251
67  .GGCCCGGCATTACCGAAGTCGGTTCGCGAAGAAATATGCGGCA 115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
251  uGlyGluAlaValValSerGlyAlaValThrProAspGluTyrIleValG 268
116  TGGCCCGCTCGATGAATAACAGAGGTGAAGCCGTCAGGCGGCAAA 165
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
268  luGlyGlyThrTrpLysIleLysGluLysValIleAlaLysGluVal 284
166  GTCTGCTTTGAACAAAAAGATCCG.....GGCGTAGTATT 203
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
285  MetValile.....ArgAsnProGluThrGlyArgGlyThrValMe 298
204  TACTGGCCCGCTCAGGCAAAATCCCGCTATTACCGCTGC..... 246
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298  tVal.....LysValAlaGlu...TyrLeuGlyProGluT 309
247  ....GAAAGCGCGTACTT.....CAGTCAGTCGTGATTCGCGTGA 285
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
309  rpValGluLysGlnValLeuThrAspGluGlnIleIleGluValAlaLys 325
286  GGCACGACCAATCGATTCGACGCTACCTACCTGAAGCGCTGCA... 333
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
326  MetGlyGlnLysIleGluAspHisTyrGlyTirpProGlnAspIleGluTr 342
334  .AAATTGAGCAGGAAAAAGTGGCGCGCAACCTGATTCAATCAGCTTAT 382
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
342  pAlaTyrAspLysaspGlyLysLeuTyrIleValGlnSerArgProI 359
383  GGACTGCGCTTCGACCCGCTCGCTTCAGCAAAATCCCTGCGGTAGATGCC 432
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
359  leThrThrLeuLysGluGluAlaThrAlaGlu...GluAlaGluGluVal 374
433  GAGCGGTGCGCATCTTCGTCAATGCGATGACACCAATCCG...CTGGC 479
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
375  GluGluAlaGluValIleLeuLysGlyLeuGlyAlaSerProGlyIleGI 391
480  TGCCGACCCCTACGCTCATCATCAAGAACGCCGCGAA.....GACTTCA 523
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
391  yAlaGlyArgValValValIlePheAspAlaSerGluIleAspLysValL 408
524  AACCGCGC...CTGTGGTATTGAGCGCGCTGACCGAACGTAATCCAT 570
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
408  ysGluGlyAspIleLeuValThrMetThrAsnProAspMetValPro 424
571  GTGTGTAAAGCAGCGGAGCGGATGCGCTCTGAAATGCT..... 612
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
425  AlaMetLysArgAlaAlaIleValThrAspGluGlyArgThrSe 441
613  ....GCCAATATCAACACATGAATTTGGCGGCCCGCATCTGCGCGCT 658
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
441  rHisAlaAlaIleValSerArgGluLeuGlyIlePro..... 453
659  TGAGTGGCAGCCACATTCATTTCATCGAGCCAGTCGCGCGCAATAAAC 708

```

```

453  ..... 453
709  GTGTGGACCATCAATATCAAGACGTGATTCGTCATCGAGCTTGTCTG 758
454  .....CysValValGlyThrLysGluAl 461
759  AACAGCGCTGTGAATACCGAGCGGTGCTTGGCGCGCTGCA... 807
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
461  aThrLysLysLeuLysThrGlyMetTyrValThrValAspGlyThrArg 478
808  ....GTCAACAAACCGCGCTCTTGGGTACCGTTTGGGTGGCGAAGGTG 852
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
478  lyLeuValTyrLysGlyIleValLysSerLeuValLysLysLysGlu 494
853  TCTCAACTTACCGCGCGGATGTTGACGCGGACCAACCGCTGATTTC 902
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
495  AlalysAlaGluGlyGlnValValAlaGlyAlaProLeuValTh 511
903  CGGTTCG.....GTATTGAACGGTG 922
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
511  rGlyThrMetValLysValAsnValSerMetProGluValAlaGluArg 528
923  CGATTGCAACAGCGCG.....CATGAT 945
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
528  laAlaAlaThrGlyAlaAspGlyValGlyLeuLeuArgAlaGluHisMet 544
946  TATTGGGACGCTACCAATCAGATTTCGTTATCGAAGAGCGCGCAG 995
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
545  lleLeuSerIleGlyGlnHisProIleLysPheIleLysGluGlyLysG 561
996  CAAAGAGCTGTTCGGTGGTGGCGCGCGGACCAATATCTCCATCA 1045
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
561  uGluLeuValGluLysLeuAlaGluGlyIleGluLysValAlaAla... 577
1046  CGCGCACCATCTCGCGCATTT.....CCTAAAA 1074
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
578  AlaPheTyrProArgProValTyrArgThrLeuAspAlaProThrA 594
1075  AACAACTCTTCAAGTT.....CACGACAGCGCTCAACGGCGCGGA 1115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
594  snGluPheArgGluMetProGlyGlyGluAspGluProGluGluArgAsn 610
1116  CGCGCCATGTGACCGCGCTTATGACGCGCTAATCGCTTGGACA 1165
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
611  ProMetLeuGlyTyrArgGly.....IleArgArgGlyLeuAspGI 624
1166  TCCTGCTACCTTCTTGGCGGATTTAATCGTCGCGATACCGACAGC 1215
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
624  nProGluLeuLeuArg...AlaGluPheLysAlaIleLysLysValValG 640
1216  GCGGAGCTTGGGTGCTTGGATTTGGACGAAGACCTCGCTTGTG 1265
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
640  luLysGlyTyrAsnAsnIleGly.....Val 648
1266  CAGCTTCGCTCG.....CCGCGCAATACGAATACGCGCGCTGT 1306
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
649  MetLeuProLeuValSerHisProGluGlnIleArgGluAlaLysArgI 665
1307  TGCAGCAAGTGTGGA 1322
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
665  eAlaArgGluValGly 670
seq_name: pir2:T45134

```

seq\_documentation\_block:  
hypothetical protein [imported] - Microbacterium ammoniaphilum (fragment)  
C;Species: Microbacterium ammoniaphilum  
C;Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
C;Accession: T45134  
R;Striebel, H.M.; Seeber, S.; Jarsch, M.; Kessler, C.  
Gene 172, 41-46, 1996  
A;Title: Cloning and characterization of the MamI restriction-modification system fro  
A;Reference number: 222923; MUID:96257250

A:Accession: T45134  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-529 <STR>  
 A:Cross-references: EMBL:X79027; NID:9984667; PIDN:CAA55649.1; PID:g1679831  
 A:Experimental source: ATCC 15354

## alignment\_scores:

Quality: 110.50 Length: 490  
 Ratio: 0.531 Gaps: 25  
 Percent Similarity: 42.449 Percent Identity: 24.694

## alignment\_block:

US-09-303-518d-131 x T45134 ..

Align seg 1/1 to: T45134 from: 1 to: 529

```

18 AGGTCTAAATCGCCATCGCGG..... 41
   |||:|||||
99 ArgAspArgAlaGluArgGlyArgGluProAlaArgProAlaAl 115:
42 .CAGACCGGAGCAATGTCGCGCATCGCCCTCGATGAAATCAAGGA 140
   |||:|||||
115 aArgAspAlaArgArgHisLeuArgLeuGlyAlaHisProAlaArgArg 132
   |||:|||||
91 TTGCTGGCGAAGATATGTCGCGCATCGCCCTCGATGAAATCAAGGA 140
   |||:|||||
132 lyAlaLeuArgArgHisArgHisArgAlaArgGlyGlu...ArgGly 147
   |||:|||||
141 AGGTGAAGCGCTCAAAAGCCCAAGTCTGTTGAAGACAAAAGAATC 190
   |||:|||||
148 ArgGlyProArgGlnGlnValPro.....ArgGlnHisProAr 160
191 CGGCGTAGTATTACTCGCGCGCTTCAGGCAAAATCGCGCTATTTCAC 240
   |||:|||||
160 gGly.....ArgArgAspArgAlaGlyArgP 169
241 CGTGGCGAAAGCGCTACTTCAGTCTGATGCTGATTCGCGTGAAGCAA 290
   |||:|||||
169 roGlyLeuHisArgArg.....ArgArgArgAlaArgArgGln 182
291 CGACCAATCGAGTTCGACGCTACGTACCTGAAGCGCTGCGCAAAATTC 340
   |||:|||||
183 GlyArgProGlnValArg.....HisGlyAspAspG 193
341 GCACGAAAGCGCGCGCAACTGATTCATTCAGGCTTATGAGCTGCG 390
   |||:|||||
193 nHisArgAlaAspProArgArgProArgAspProArg.....AlaG 207
391 CTTCGCGACCGCTCGTTCAGCAAAATCCCTGCGTAGATCGCGACCGTT 440
   |||:|||||
207 lyHisHisProLeuArgGluGlyGlnGlyGlnGluAlaArgProLeu 223
441 CGCATCTTCGTCATTCGATGACACCAATCCGCTGCGTGGCGCGACCTA 490
   |||:|||||
224 Arg.....AlaArgGlyGlnGlyGlyProGlyGlyProAl 236
491 CGGTCTCATCT.....CAAAGAGCGCGCGCAAGATTC 522
   |||:|||||
236 aGlyArgHisProAlaLeuGlyAspGluAspArgProArgAlaArgGlnG 253
523 AAACGGCGCTGTTGTTGTTAGCGCGCTGACCGAAGCTAAATCCATGT 572
   |||:|||||
253 lyAlaArg..... 255
573 GTCTAAGCAGCGCGCAGAGCTGCGCTGTAATAATCGCCATATCG 622
   |||:|||||
256 .....AlaAlaHisArgGluAlaLeuHisLeuArgLeuGlnArgAr 269
623 AAACACATGAATTTGGCGCGCGCATCTCGCGGCTTGAAGTGGACCGAC 672
   |||:|||||
269 gArg.....GlyArgAlaAspGlyCysArgAlaGlnGlyArgAla 283

```

```

673 ATTCAATTCATCGAGCCAGT.....CGCGCGCAATAAAC 707
   |||:|||||
283 rgArgAlaArgArgAlaGlyGlyGlyLeuProArgArgGlu..... 297
708 CGTGTGGACCATCAATTCAAGACGCTGATTGCTATCGGACGTTTGTTCG 757
   |||:|||||
298 .....AspArgValArgAlaHisArgProArgProAr 308
758 TAACAGCGCGCTCTGAATAC.....CGAGCGCTGTTGCTTGGCGG 798
   |||:|||||
308 gGlyArgGlyGlyAlaAlaArgLeuAspArgAlaGlyValGlyAla 325
799 GGCCTGCA..... 806
   |||:|||||
325 rgProAlaArgProArgArgLeuArgHisProArgProAlaAspLeuPro 341
807 .....AGTCAACAAACGCGCCTCTTCGCTACCGTTTGGTGGCGA 847
   |||:|||||
342 HisGlyGlyProGlnGlyValAlaArgLeuAspHisProGlnGlyLeuG 358
848 AGGTGTCCTCAACTTACCGCGCGCAATTGCTGACGCGGACACCGCGTG 897
   |||:|||||
358 uGlyAlaAlaGlySerArgArg.....HisProHis 369
898 ATTTCCGCTCGGTATTGAA.....CGTGCATTCACACAGCGCGCGCA 941
   |||:|||||
369 rgLeuArgAlaArgLeuHisGlnGlyArgGlyAspLeuLeuArgArg... 384
942 TGATTATTTGGGACGTACCAATCAGATTTCGTTATCGAAGAAGCGC 991
   |||:|||||
385 .....ProArgArgAspArgLeuGlyArgArgGlyPr 395
992 GCACAAAGAGCTGTCGCGTGGCTGCGCGCGCGCGCAACAACTCTCAAGTT 1041
   |||:|||||
395 oArgGlnGlyGln.....GlyAlaHisGlyGlyGlnGlyLeuA 408
1042 ATCAGCGCGCACCTCTCGCGCTGCTTCCCTAAATAAAACAACTCTCAAGTT 1091
   |||:|||||
408 rgHisAlaGlyArgArgArg.....ArgGlyValProSerGlnLeuAlaVal 424
1092 CACGACAG.....CCGTCAACGCGCGCGCGCGCGCGCGCGCGCGCA 1132
   |||:|||||
424 euArgGlnGlyValCysGlnAlaSerAlaThrLeuAlaLeuTrpMetThr 440
1133 TCGSCATTTATGAGCGCGTAATCCCTTGGACATCTGCTGCTTCACTGCTT 1182
   |||:|||||
441 SerGlyArgLeu.....LeuArgGluLeuAlaGlyLeuLeuCysAr 454
1183 TTGCGCATTTAATCTCGCGCGCATACCGACA..... 1213
   |||:|||||
454 gAspSerValGluSerHisLeuAspValThrThrTrpArgAlaAspAla 471
1214 .....CGCGCGCGCGCTTGGTGTGTTGGAATTTGGAGAAAGAACCTCG 1258
   |||:|||||
471 erileSerArgArgThrTrpThrValMetAsn...ThrArgLeuArgGln 486
1259 CTTTGTGCGAGTCTGCTCGCGCGCGCAATACGAATACGCGCGCGCTGTTG 1308
   |||:|||||
487 AlaGlnLysSerLeuAlaAlaArgArgProThrThrSerSerSerIleSe 503
1309 CGCAAAAGTCTGGAAA 1324
   |||:|||||
503 rArgSerCysTrpGln 508

```

seq\_name: pir2:T34434

seq\_documentation\_block:

hypochemical protein K06A9.1a - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 18-Feb-2000

C/Accession: T34434

R:Geisel, C.; Gattung, S.



C:Species: Rhodobacter sphaeroides  
C:date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 28-Jul-2000  
C:Accession: T50711  
R:Choudhary, M.; Kaplan, S.  
Nucleic Acids Res. 28, 862-867, 2000

121 CCTCGATGAAATCAAGGAAGTCAAGCCGTCAAAAAGCCGCAAGTGCT 170  
|||||: |||: ||| ||||| |||: |||: |||  
116 ProGlyThrGluIleAlaGlyGlu.....GlyArgIleLeu 128  
171 G.....T 172  
|  
128 uThrAlaGlyGlyMetAspAlaHisIleHisPheIleCysProGlnGlnI 145  
173 TTGAAGACAAAAGAAATCCGGCGCTAGTATT.....ACT 207  
:|||||: |||: |||: |||: |||: |||  
145 leGluAspSerLeuHisSerGlyIleThrMetLeuGlyGlyGlyThr 161  
208 GCGCCGGCTTCAGGCAGAAATCCGCGCT.....AT 236  
:|||||: ||| |||: |||: |||: |||: |||  
162 GlyProAlaHisGlyThrLeuAlaIleThrThrLeuHisAlaGlyAlaLeuAl 178  
237 TCACCGT.....GCGCAAAAGCGCGTACTTCAGTCAGTCGTA 274  
||||| ||| |||: |||: |||: |||: |||: |||  
178 aHisArgAlaAspAlaAlaGlyArgLeuProAspGlnProArgL 195  
275 TTGCCGTTGAAGGCAACGACGAATCCAGTTCGA..... 309  
: |||: |||: |||: |||: |||: |||: |||: |||  
195 euArgGlyGlnGlyCysGlnProAlaArgAlaArgGlyAlaGly 211  
310 .....CGTACGTCACCTGAAGCGCTGGCAAAATGAGCAGCGCAAA 350  
||| ||| ||| ||| ||| ||| |||: |||: |||: |||  
212 ProArgGlyArgLeuValProGluAlaAlaArgGlyLeuGlyHisHisAl 228  
351 AGTGC CGCGCAACCTGATTCANTCAGCCTT.....ATGACTGCGCTT 393  
: |||: |||: |||: |||: ||| ||| |||: |||: |||: |||  
228 acysGlyHisArgLeuLeuProValGlyArgProHisGlyCysAlaG 245  
394 CG.....CACCGTCGCTTCAGCAAAATCCCTGCGCTAGATGCCGA 434  
: ||| ||| ||| ||| ||| ||| |||: |||: |||: |||  
245 lyAspProHisArgHisAlaGluArgGluArgLeuArgGluHis 261  
435 GCGGTTCCCATCTTCGT.....CAATCGATGCACACCAATCCGC 475  
||| ||| ||| ||| ||| ||| |||: |||: |||: |||  
262 AlaGlyArgHisArgArgAlaHisAspProCysValProHisArgGlyCy 278  
476 TGGTCCGCGACCTTACGGTCATCATCAAGACGCGCGAAGACTTCAAA 525  
: ||| ||| ||| |||: |||: |||: |||: |||: |||  
278 scLyArgArgProCysAlaArgHisGlnGlyGlyArgArgGluA 295  
526 CCGCGCCTGTT..... 536  
: |||: |||: |||: |||: |||: |||: |||: |||  
295 rgHisProValLeuAspGlnSerHisHisAlaLeuHisArgGlnTyrgly 311  
537 .....GTTATTGAGCGCCTGACCGCAAGCTA 562

